صحافة البيانات

كيف نستخرج الأخبار

من أكوام الأرقام والمعلومات في الإنترنت

صدر هذا الكتاب بلغات عدة،
وترجم إلى العربية عن الأصل الإنجليزي بعنوان:

Data Journalism

Edited by Liliana Bournegru, Lucy Chambers and Jonathan Gray

الكتاب بالإنجليزية متاح على الإنترنت، ومتاح النقل عنه
وأصدره بالعربية بالاتفاق مع الناشرين
قطاع ضبط الجودة والمعايير التحريرية بشبكة الجزيرة.

صحافة البيانات

كيف نستخرج الأخبار

من أكوام الأرقام والمعلومات في الإنترنت

المحررون:

جوناثان غراي، ليليان بونيغرو،

لوسي تشيمبرز

ترجمة: ندى رمضان





بسم الله الرحمن الرحيم

الطبعة الأولى

2015 م – 1437 هـ

ردمك 978-614-01-1744-0



ضبط الجودة والمعايير التحريرية
QUALITY ASSURANCE & EDITORIAL STANDARDS
شبكة الجزيرة الإعلامية – الدوحة – قطر
هاتف: 44896744 (974+)
فاكس: 44896888 (974+)

الدار العربية للعلوم ناشرون ش.م.ل
Arab Scientific Publishers, Inc. S.A.L

عين التينة، شارع المفتي توفيق خالد، بناية الريم
هاتف: 786233 – 785108 – 785107 (1-961+)
ص.ب: 5574-13 شوران – بيروت 2050-1102 – لبنان
فاكس: 786230 (1-961+) – البريد الإلكتروني: asp@asp.com.lb
الموقع على شبكة الإنترنت: http://www.asp.com.lb

تصميم الغلاف: **علي القهوجي**

التنضيد وفرز الألوان: **أبجد غرافيكس**، بيروت – هاتف 785107 (1-961+)
الطباعة: **مطابع الدار العربية للعلوم**، بيروت – هاتف 786233 (1-961+)

المحتويات

الفصل الأول
مقدمة

الفصل الثاني
في غرفة الأخبار

الفصل الثالث
دراسات لحالات نموذجية

الفصل الرابع
الحصول على البيانات

الفصل الخامس
فهم البيانات

الفصل السادس
تقديم البيانات

مقدمة الترجمة العربية

ولدت هذه الترجمة في مخاض طويل، فقد أنجزتها السيدة ندى رمضان مع إشراف من الزميل أحمد القوني، الذي ساهم أيضاً بقدر من الترجمة، وتجشما في سبيل الخروج بمقابل عربي للعديد من المصطلحات الكثير من العنت، على أنه تم الإبقاء على عدد من المصطلحات في صورتها الأصلية لأنها أصبحت دوال عالمية على مفاهيم وعمليات بعينها. وانتقلت الترجمة بعد ذلك إلى قطاع ضبط الجودة حيث راجعها الزميل محمد شقير.

وكان للزملاء في الجزيرة نت نظرة فاحصة في هذه الترجمة. وقد اجتمعت هذه الجهود لإخراج هذا الكتاب. على أن الأشهر امتدت قبل التمكن من إخراجه، واضطررنا في المرحلة الأخيرة إلى الإسراع في إرساله إلى المطبعة، مما أوقعنا في بعض الخلط فيما يتعلق بأسماء المشاركين فيه ولا سيما فيما يتعلق بالترجمة الأساسية التي أنجزتها، مع ضبط المصطلحات وتعريب الجزء الأكبر منها، السيدة ندى رمضان.

نرجو أن يكون في هذا الكتاب ما يفيد الصحفي المقبل على حقبة جديدة؛ وإن كل صحفي يتعامل مع شؤون أوسع نطاقاً من منطقته الضيقة مقبل على هذه الحقبة التي تتميز بتوفر المعلومات والأرقام على نحو غير مسبوق. فمن هذا الكم الهائل من المعلومات يمكن استخراج الكثير من الأخبار التي كانت في الماضي تظل حبيسة الجداول والبيانات في أدراج الدوائر الحكومية والمؤسسات المختلفة.

ضبط الجودة والمعايير التحريرية

شبكة الجزيرة

الدوحة 25 نوفمبر / تشرين الثاني 2015

كلمة هيئة تحرير الكتاب

جوناثان غراي – ليليان بونيغرو – لوسي تشيمبرز

ولد كتاب "صحافة البيانات" في ورشة بلندن عام 2011. ومنذ ذلك الحين انتشر هذا اللون من الاستقصاء الصحافي، وسرعان ما اتسع نطاقه ليصبح جهدا دولياً جماعياً يعمل فيه عشرات الصحافيين المؤمنين بهذا النمط الصحافي الواعد. وفي غضون ستة أشهر منذ البدء في وضع الكتاب وحتى صدور أول طبعة منه، ساهم المئات – منهم من نعرف ومنهم من لا نعرف– بأشكال وطرق شتى في تشكيله. وحاولنا أن ندرج مساهمات العدد الأكبر منهم ضمن هذا الكتاب. لهؤلاء جميعا ولمن لم يذكر اسمه أو تدرج مساهماته ضمن الكتاب نقدم الشكر، وندعوهم لمواصلة العمل.

قائمة المساهمين:

القائمة التالية تضم أسماء من ساهموا بشكل مباشر أو غير مباشر في هذا الكتاب بطبعته الحالية.

قام بتصميم الأشكال البيانية والصور "كيت هدسون":

- Gregor Aisch, Open Knowledge Foundation
- Brigitte Alfter, Journalismfund.eu
- David Anderton, Freelance Journalist
- James Ball, The Guardian
- Caelainn Barr, Citywire
- Mariana Berruezo, Hacks/Hackers Buenos Aires
- Michael Blastland, Freelance Journalist
- Mariano Blejman, Hacks/Hackers Buenos Aires
- John Bones, Verdens Gang
- Marianne Bouchart, Bloomberg News
- Liliana Bounegru, European Journalism Centre

- Brian Boyer, Chicago Tribune
- Paul Bradshaw, Birmingham City University
- Wendy Carlisle, Australian Broadcasting Corporation
- Lucy Chambers, Open Knowledge Foundation
- Sarah Cohen, Duke University
- Alastair Dant, The Guardian
- Helen Darbishire, Access Info Europe
- Chase Davis, Center for Investigative Reporting
- Steve Doig, Walter Cronkite School of Journalism of Arizona State University
- Lisa Evans, The Guardian
- Tom Fries, Bertelsmann Stiftung
- Duncan Geere, Wired UK
- Jack Gillum, Associated Press
- Jonathan Gray, Open Knowledge Foundation
- Alex Howard, O'Reilly Media
- Bella Hurrell, BBC
- Nicolas Kayser–Bril, Journalism++
- John Keefe, WNYC
- Scott Klein, ProPublica
- Alexandre Léchenet, Le Monde
- Mark Lee Hunter, INSEAD
- Andrew Leimdorfer, BBC
- Friedrich Lindenberg, Open Knowledge Foundation
- Mike Linksvayer, Creative Commons
- Mirko Lorenz, Deutsche Welle
- Esa Mäkinen, Helsingin Sanomat
- Pedro Markun, Transparência Hacker
- Isao Matsunami, Tokyo Shimbun
- Lorenz Matzat, OpenDataCity
- Geoff McGhee, Stanford University
- Philip Meyer, Professor Emeritus, University of North Carolina at Chapel Hill
- Claire Miller, WalesOnline
- Cynthia O'Murchu, Financial Times

صحافة البيانات

كيف نستخرج الأخبار

من أكوام الأرقام والمعلومات في الإنترنت

الفَصْـلُ الأوّلُ

مُقَـدِّمَـة



ما هي صحافة البيانات؟ ما هي إمكانياتها؟ ما هي حدودها؟ ما هو مصدرها؟ في هذا الباب سننظر في ماهية صحافة البيانات وفيما قد تعنيه بالنسبة لمؤسسات الأخبار. يتحدث كل من بول برادشو (جامعة برمنغهام) وميركو لورينز (إذاعة دويتشه فيله) قليلاً عما يميز صحافة البيانات. كما يخبرنا كبار الصحافيين لماذا يعتبرونها هامة وما هي أمثلتهم المفضلة. وأخيراً، تضع ليليانا بونجيرو (مركز الصحافة الأوروبي) صحافة البيانات في سياقها التاريخي الأوسع.

ما هي صحافة البيانات؟

ما هي صحافة البيانات؟ يمكنني أن أجيب بكل بساطة بأنها الصحافة التي تصنع باستخدام البيانات. ولكن هذا قد لا يقدم الجواب الشافي، ولا يساعد كثيراً.

إن كلا المصطلحين 'صحافة' و'بيانات' يحمل في طياته قدرا من التعقيد، إذ ينظر البعض إلى 'البيانات' باعتبارها مجموعة من الأرقام غالباً ما تكون مجمعة في جداول بيانية. وقبل 20 عاماً، كان هذا هو النوع الوحيد تقريباً من البيانات التي يتعامل معها الصحافيون. ولكننا اليوم نعيش في عالم رقمي، يمكن وصف أي شيء وكل شيء فيه تقريباً بالأرقام، وهو ما يحدث بالفعل.

من الأمثلة الصارخة على ذلك، تاريخك المهني مثلا، إضافة إلى 300,000 وثيقة سرية أخرى، وجميع من هم في دائرة أصدقائك ومعارفك؛ يمكن وصف كل هذا الكم الضخم من البيانات باستخدام رقمين فقط: الأصفار والآحاد. فالصور ومقاطع الفيديو والصوت توصف جميعا بنفس الرقمين: الأصفار والآحاد. كذلك فإن جرائم القتل والأمراض والأصوات الانتخابية والفساد والأكاذيب كلها تختزل إلى: أصفار وآحاد.

ما الذي يجعل صحافة البيانات تختلف عن غيرها من أنواع الصحافة وتصنيفاتها المختلفة؟ ربما الاختلاف في الكم الضخم من الإمكانيات الجديدة والبدائل التي تصبح متاحة حين يجمع المرء بين «الفضول التقليدي للصحافي» والقدرة على سرد قصة مؤثرة هي موضوع خبر ما، ضمن هذا النطاق الهائل من البيانات الرقمية المتاحة في متناول أيديكم.

ويمكن لتلك الإمكانيات أن تأتي في أي مرحلة من عملية ممارسة الصحافة: عند استخدام البرمجة لتحويل عملية جمع المعلومات ومزجها من مصادر كالحكومة المحلية والشرطة والمصادر المدنية الأخرى إلى عملية دمج تلقائية، كما فعل «آدريان هولوفاتي» في قسم الجرائم من صحيفة شيكاغو تريبيون المسمى Chicago Crime (http://chicago.everyblock.com/crime/) ثم في

موقع Every Block (http://www.everyblock.com/).

أو باستخدام برنامج كمبيوتر لإيجاد روابط بين مئات الآلاف من الوثائق، كما فعلت صحيفة التليغراف في موضوع نفقات النواب (http://tgr.phs/mps-expenses/).

يمكن لصحافة البيانات أن تساعد الصحافي على سرد خبر معقد عبر طرح يتضمن إدخال الرسوم البيانية، مثلما فعل «هانز روزلينج» في مناقشاته الرائعة حول توضيح الفقر العالمي مرئياً باستخدام تطبيق Gapminder (http://www.gapminder.org/) الذي أثار ملايين الآراء عالمياً. والعمل الشهير «لدايفيد مكاندليس» الذي استخرج فيه أرقاما كبيرة - مثل وضع الإنفاق العام في السياق المطلوب لإيضاح أهمية القضية، أو التلوث الناتج عن بركان آيسلندا الذي تمت الوقاية منه - والذي يوضح أهمية وجود تصميم واضح في موقع Information is Beautiful (http://www.informationisbeautiful.net/).

Investigate your MP's expenses

Join us in digging through the documents of MPs' expenses to identify individual claims, or documents that you think merit further investigation. You can work through your own MP's expenses, or just hit the button below to start reviewing. (Update, Fri pm: we now have a virtually complete set of expenses documents so you should be able to find your MP's) Already created an account? Log in here.

We have **458,832** pages of documents. **32,755** of you have reviewed **225,443** of them. Only **233,389** to go...

Start reviewing

Please read our privacy policy to find out how we use your data. You must also read our terms of service. By reviewing pages, you are agreeing that you have read the terms of service, and that you agree to them.

نموذج 1-1. حقق في نفقات نائبك في البرلمان (الغارديان)

وبإمكان صحافة البيانات شرح علاقة الخبر بالفرد، مثلما تفعل شبكة بي. بي. سي وفاينانشال تايمز بانتظام في العرض التفاعلي للميزانية (حيث يمكنك معرفة تأثير الميزانية عليك أنت، بدلاً من فلان أو علان). كما يمكنها أن تتيح

عملية جمع الأخبار نفسها، مثلما تفعل صحيفة الغارديان بشكل ناجح عبر مشاركة البيانات والسياق والأسئلة مع مدونة البيانات الخاصة بكل ذلك (http://www.guardian.co.uk/news/datablog).

يمكن للبيانات أن تكون مصدراً لصحافة البيانات وأن تكون الأداة التي يتم عن طريقها سرد الخبر، أو أن تكون كليهما. ومثل أي مصدر آخر، ينبغي أن تعامل بشيء من الشك؛ ومثل أي أداة، ينبغي أن ندرك كيف يمكنها تشكيل، وأيضاً تقييد، الأخبار التي تصنعها.

--- بول برادشو، جامعة برمنغهام

لماذا يتعين على الصحافيين استخدام البيانات

تتعرض الصحافة لنوع من الحصار. ففيما مضى، كنا نحن – الصحافيين – الوحيدين الذين يستخدمون التكنولوجيا لنسخ ما حدث في الليلة السابقة وتوزيعه. وكانت المطبعة بمثابة بوابة، فإذا أراد أي شخص الوصول إلى أناس في مدينة أو منطقة ما في الصباح التالي، لجأ إلى الصحف. ثم أمسى هذا في خبر كان.

اليوم تتدفق الأخبار أثناء حدوثها من عدة مصادر ومن شهود ومدونات، وما يحدث تتم تصفيته من خلال شبكة هائلة من الروابط الاجتماعية، كما يتم تصنيفه والتعليق عليه، وفي كثير من الأحيان يتم تجاهله.

ولهذا السبب تزداد أهمية صحافة البيانات، حيث تتزايد أهمية الجمع والانتقاء وتخيل ما يحدث بشكل يتجاوز محدودية ما تستطيع العين رؤيته. ففي الاقتصاد العالمي الحالي، توجد روابط غير مرئية بين منتجات مثل عصير البرتقال الذي تشربه في الصباح والقهوة التي تصنعها، وبينك أنت والآخرين. لغة هذه الشبكة هي البيانات: نقاط صغيرة من المعلومات التي قد لا تكون مهمة بشكلها منفردة، ولكنها تصبح في غاية الأهمية حين يُنظر إليها من الزاوية المناسبة.

هناك الآن بعض الصحافيين الرواد الذين يوضحون بالفعل كيف يمكن استخدام البيانات للتوصّل إلى رؤية أعمق بشأن ما يحدث حولنا وكيف يمكن لذلك أن يؤثر علينا.

يمكن لتحليل البيانات أن يكشف عن 'شكل الخبر' (سارا كوهين)، أو أن

يزودنا بـ ’كاميرا جديدة‘ (دايفيد مكاندليس). وباستخدام البيانات، ينتقل التركيز الرئيسي لمهمة الصحافيين من كونهم أول من ينقل الخبر إلى كونهم يبلغوننا بما يعنيه بالفعل تطور ما لحدث معين. يمكن أن تكون المواضيع متباعدة ومتسعة النطاق. قد يكون الموضوع: الأزمة المالية القادمة في طور التشكل حاليا، أو الأبعاد الاقتصادية وراء المنتجات التي نستهلكها، أو سوء استخدام الأموال، أو الأخطاء السياسية؛ تُقدّم هذه الموضوعات كلها في عرض مرئي مُقنع لا يدع مجالاً واسعا للجدال بشأنها.

لهذا يجب على الصحافيين النظر إلى البيانات باعتبارها فرصة. فعلى سبيل المثال، يمكنهم أن يوضحوا كيف سيؤثر مفهوم مجرد من قبيل (البطالة) على الناس وفقاً لأعمارهم أو جنسهم أو مستوى تعليمهم. فاستخدام البيانات يحول أمراً مجرداً إلى أمر يستطيع الجميع فهمه والارتباط به ذهنيا.

تستطيع البيانات أن تخلق أدوات حساب شخصية لمساعدة الناس على اتخاذ القرارات سواء أكان القرار بخصوص شراء سيارة أو منزل، أم الاختيار ما بين التعليم الأكاديمي أو المهني، أم التحقق من التكاليف لتلافي الديون.

كما تستطيع البيانات أن تحلل ديناميكيات المواقف المعقدة مثل أحداث الشغب أو النقاشات السياسية، وأن تظهر العيوب وتساعد الجميع على رؤية الحلول المتاحة للمشاكل المعقدة.

إن معرفة كيفية البحث في البيانات وتنقيتها وتوضيحها على نحو مرئي تحدث تحولا كبيراً في مهنة جمع المعلومات أيضاً. فالصحافيون الذين يتمكنون منها سيجدون أن إرساء المقالات على أسس من الحقائق والأفكار أمر مريح. فهذا يعني قدراً أقل من التخمين والبحث عن اقتباسات – وبدلاً من ذلك، يمكن للصحافي أن يتخذ موقفاً قوياً مدعوما بالبيانات، وقد يكون لذلك تأثير كبير على دور الصحافة.

إضافة إلى ذلك، فإن للدخول إلى مجال صحافة البيانات بُعداً مستقبلياً. فاليوم، حين تقلل غرفة الأخبار من عدد الموظفين، يأمل معظم الصحافيين أن ينتقلوا إلى العلاقات العامة. أما صحافيو البيانات، أو علماء البيانات، فهُم مجموعة مرغوبة، ليس فقط في مجال الإعلام: فالشركات والمؤسسات حول العالم تبحث

عن ’صُنّاع المعنى‘ وعن المتخصصين الذين يمكنهم التنقيب في البيانات وتحويلها إلى شيء ملموس.

البيانات مجال واعد، وهذا ما يثير حماس غرف الأخبار ويجعلها تبحث عن نوع جديد من الصحافيين. فتوافر إجادة العمل بالبيانات بالنسبة للصحافيين المستقلين يُعتبر طريقاً إلى عروض جديدة وإلى دخل ثابت أيضاً. انظر إلى الأمر هكذا: بدلاً من تعيين صحافيين يسارعون إلى تعبئة الصفحات والمواقع الإلكترونية بمحتوى منخفض القيمة، يمكن لاستخدام البيانات أن يخلق طلباً على الباقات التفاعلية، حيث قد تكون الطريقة الوحيدة للعمل هي قضاء أسبوع كامل سعيا للإجابة عن سؤال واحد. هذا التغيير هو موضع ترحاب في عدة مجالات إعلامية.

يوجد عائق واحد أمام الصحافيين يمنعهم من استغلال هذه الإمكانية: التدريب بهدف تعلم كيفية العمل باستخدام البيانات وإجادته، مروراً بكافة الخطوات، بدءاً من السؤال الأول البسيط، وصولاً إلى سبق صحفي كبير مبني على البيانات.

يشبه العمل باستخدام البيانات الدخول إلى أرض واسعة مجهولة. فللوهلة الأولى، تحيّر البيانات الخام، في صورتها البدائية المجردة، العيون والعقول. إذن فبهذا الشكل تكون البيانات مربكة وغير عملية، ويصعب للغاية صياغتها بشكل صحيح لرؤية ما تعنيه، فهي تحتاج إلى صحافيين ذوي خبرة ولديهم جلد يمكنهم من النظر في مادة خام عادة ما تكون مربكة ومملة، وتمكنهم الخبرة أيضاً من ’رؤية‘ الأخبار المتوارية بداخل البيانات.

الدراسة المسحية

أجرى مركز الصحافة الأوروبي دراسة مسحية (http://bit.ly/ddjnet-survey) لمعرفة المزيد حول احتياجات الصحافيين من التدريب. وقد وجدنا أن هناك استعداداً كبيراً للخروج عن إطار الراحة المعتاد في الصحافة التقليدية، واستثمار الوقت لاكتساب وإجادة مهارات جديدة. أظهرت نتائج الدراسة أن الصحافيين يرون الفرصة، ولكنهم يحتاجون إلى بعض الدعم ليتجاوزوا المشاكل الأولية التي قد تعيقهم عن العمل باستخدام البيانات. هناك يقين بأنه إذا تم تبني

صحافة البيانات عالمياً، فسرعان ما سيتحسن سير العمل والأدوات والنتائج. ويواصل الرواد مثل صحف الغارديان والنيويورك تايمز وتكساس تريبيون ودي تسايت رفع معايير أخبارهم القائمة على البيانات.

هل ستبقى صحافة البيانات حكراً على بضعة رواد؟ أم ستحصل كل مؤسسة اخبارية قريباً على فريقها الخاص لصحافة البيانات؟ نأمل أن يساعد هذا الدليل المزيد من الصحافيين وغرف الأخبار على استغلال هذا المجال الصاعد.

--- ميركو لورينز، إذاعة دويتشه فيله



نموذج 2–1. دراسة مركز الصحافة الأوروبي حول احتياجات التدريب (العنوان: 108 استجابة من 40 بلداً)

ما أهمية صحافة البيانات؟

سألنا بعض أبرز ممارسي صحافة البيانات ومؤيديها عن سبب اعتقادهم بأنها تطور هام، وهذا ما قالوه:

تصفية تدفق البيانات

عندما كانت المعلومات نادرة، كنا نكرس معظم جهودنا للتقصي والجمع. والآن مع وفرة المعلومات، فإن ما أصبح أكثر أهمية هو معالجتها. نعالج المعلومات على مستويين: (1) التحليل: لاستخراج مضمون له مغزى وبناء له شكل محدد أو هيكل من التدفق المتواصل للبيانات، و(2) العرض: لإيصال كل ما هو هام وذي صلة إلى المستهلك. وبنفس الأسلوب العلمي، فإن صحافة البيانات تفصح عن أساليبها وتعرض استنتاجاتها بطريقة تتيح التحقق منها عن طريق النسخ المتماثل.

— فيليب ماير، أستاذ فخري بجامعة كارولينا الشمالية في تشابل هيل

أساليب جديدة لسرد الأخبار

'صحافة البيانات' مصطلح شامل يضم مجموعة متنامية من الأدوات والأساليب والمناهج لرواية الأخبار. وقد تتضمن أي شيء، بدءاً من جمع الأخبار بالطريقة التقليدية بمساعدة الكمبيوتر (باستخدام البيانات كـ 'مصدر') وصولاً إلى أكثر تطبيقات الأخبار والتوضيح المرئي تطوراً. والهدف الجمعي صحافي: توفير المعلومات والتحليلات لتعريفنا جميعاً بأهم القضايا والأحداث اليومية.

— آرون بيلهوفر، صحيفة نيويورك تايمز

تشبه الصحافة المصورة ولكن باستخدام الكمبيوتر المحمول

تختلف 'صحافة البيانات' عن 'صحافة الكلمات' فقط في أنها تستخدم أدوات مختلفة. فنحن جميعنا نعمل في التقصّي ونقل الأخبار والربط بين الأخبار. وهذا يشبه 'الصحافة المصورة'، لكن فقط باستخدام الكمبيوتر المحمول بدلا من الكاميرا.

— بريان بوير، صحيفة شيكاغو تريبيون

صحافة البيانات هي المستقبل

ينظر إلى الصحافة المستندة إلى البيانات على أنها المستقبل، لذا يجب على

الصحافيين أن يكونوا ماهرين في توظيف البيانات. فيما مضى، كنت تستطيع الحصول على الأخبار بالتحدث مع الناس في الحانات، وربما لا يزال بإمكانك أن تفعل ذلك أحياناً، أما الآن فينبغي عليك الدخول في عمق البيانات وتجهيز نفسك بالأدوات اللازمة لتحليلها وانتقاء ما يثير الاهتمام منها. ومن خلال الإبقاء عليها ضمن إطار واحد، ستتمكن من مساعدة الناس على رؤية الأمور من منظور منطقي متماسك، ومعرفة ما يحدث في البلاد.

—— تيم بيرنرز – لي، مؤسس شبكة الإنترنت العالمية

تحليل الأرقام وصياغة الكلام

تضيق صحافة البيانات الفجوة بين المتخصصين في مجال الإحصاء وصناع الكلمات. فتقوم بتحديد القيم المتطرفة والتعرف على التوجهات الهامة ليس فقط على مستوى الإحصاءات بل الأمور ذات الصلة أيضاً لفك طلاسم عالم اليوم المعقد بطبيعته.

—— دايفيد أندرتون، صحفي مستقل

تحديث مجموعة مهاراتك

صحافة البيانات مجموعة جديدة من المهارات للبحث في المصادر الرقمية وفهمها وتوضيحها مرئياً، في وقت لم تعد فيه المهارات الأساسية للصحافة التقليدية كافية. وهي ليست بديلاً عن الصحافة التقليدية، بل إضافة إليها.

في الوقت الذي تتحول فيه المصادر إلى مصادر رقمية، يمكن للصحافيين – بل ويجب عليهم – أن يكونوا أقرب إلى تلك المصادر. فقد أتاحت شبكة الإنترنت إمكانيات تتجاوز استيعابنا حالياً. كما أن صحافة البيانات هي مجرد بداية تطور ممارساتنا الماضية للتكيف مع المحتوى الإلكتروني.

لصحافة البيانات غرضان هامان بالنسبة لمؤسسات الأخبار: العثور على أخبار فريدة (ليست من وكالات الأنباء) وتنفيذ مهمة المراقبة. وخاصة في أوقات الخطر المالي، حين تهدف الصحف إلى تحقيق هذه الأهداف الهامة.

تّعتبر صحافة البيانات شديدة الأهمية بالنسبة للصحف المحلية. ولدينا مقولة: ’بلاطة مخلوعة أمام بابك خبر أهم من أحداث الشغب في بلد بعيد’. فهي أمام ناظريك، وتؤثر على حياتك بشكل مباشر. وفي الوقت ذاته انتشر التحويل الرقمي في كل مكان. ولأن الصحف المحلية تحدث ذلك التأثير المباشر على أحيائها السكنية ويتم تحويل المصادر رقمياً، ينبغي على الصحفي معرفة كيفية إيجاد الخبر في البيانات وتحليله وعرضه على نحو يمكن أن يتصوره القارئ بوضوح.

— جيري فيرمانين، موقع NU.nl

حل لتفاوت المعلومات

إن تفاوت المعلومات – لا أقصد نقص المعلومات، بل عدم القدرة على استيعابها ومعالجتها بالسرعة والحجم اللازمين – يعتبر من أكبر المشاكل التي يواجهها المواطنون عند اتخاذ القرارات بشأن حياتهم. فالمعلومات المأخوذة من الإعلام المطبوع والمرئي والسمعي تؤثر على قراراتهم وأفعالهم. وصحافة البيانات الجيدة تساعد على مواجهة مشكلة تفاوت المعلومات.

— توم فرايز، مؤسسة بيرتلسمان

حل لمشاكل العلاقات العامة المبنية على البيانات

جعل توفر أدوات القياس بأسعار منخفضة – مع فعاليتها وشمولها جميع جوانب المجتمع – صناع القرار يقبلون على قياس مقدار تقدم سياساتهم ومراقبة التوجهات وتحديد الفرص.

فالشركات ماضية في التوصّل إلى مقاييس جديدة لقياس جودة أدائها. ويحب السياسيون التفاخر بالانخفاض في معدل البطالة والزيادة في إجمالي الناتج المحلي. بينما كان انعدام الرؤى الصحفية العميقة في قضايا مثل إنرون Enron ووورلدكوم Worldcom ومادوف Madoff وسوليندرا Solyndra يدل على عجز العديد من الصحافيين عن فهم مؤشرات الأرقام بشكل واضح. فالأرقام غالباً تُقبل على ظاهرها بدلاً من استنتاج الحقائق، فالأرقام توحي بالجدية حتى وإن كانت ملفقة بالكامل.

ستساعد إجادة التعامل مع البيانات الصحافيين على شحذ حسهم النقدي عند التعامل مع الأرقام، كما يؤمل أن تساعدهم على استعادة بعض من النفوذ في تعاملهم مع أقسام العلاقات العامة.

—— نيكولاس كايسر – بريل، شبكة Journalism++

توفير تفسيرات مستقلة للمعلومات الرسمية

بعد الزلزال المدمر وكارثة المفاعل النووي في فوكوشيما عام 2011، تأكد إعلاميو اليابان من أهمية صحافة البيانات، واليابان بشكل عام متأخرة في مجال الصحافة الرقمية.

لقد حيرنا انعدام البيانات الموثوق بها لدى الحكومة والخبراء. فعندما أخفى المسؤولون عن الجمهور بيانات SPEEDI (التنبؤ بانتشار المواد المشعة)، لم نكن قادرين على فك رموزها حتى وإن تسربت. بدأ المتطوعون في جمع بيانات النشاط الإشعاعي مستخدمين أدواتهم الخاصة، ولكننا لم نكن مجهزين بمعرفة الإحصائيات أو التدخل أو التوضيح المرئي وما إلى ذلك. ينبغي على الصحافيين أن يتمكنوا من الحصول على البيانات الخام الأولية، وأن يتعلموا ألا يعتمدوا على التفسيرات الرسمية لها.

—— إساو ماتسونامي، صحيفة طوكيو شيمبون

التعامل مع سيل البيانات

تواصل التحديات والفرص التي تقدمها الثورة الرقمية التشويش على الصحافة. ففي عصر وفرة المعلومات، يحتاج كل من الصحافيين والمواطنين على حد سواء إلى أدوات أفضل، سواء أكنا نروج منشورات حركة ما في الشرق الأوسط في القرن الحادي والعشرين، أو نعالج مجموعة من البيانات في وقت متأخر، أو نبحث عن أفضل طريقة لتوضيح أهمية المياه بشكل مرئي لجمهور من المستهلكين. وبينما نصارع تحديات الاستهلاك التي يفرضها هذا السيل من البيانات، ظهرت منصات جديدة للنشر ومكَّنت الجميع من التجمع والتشارك في البيانات رقمياً، محولين إياها إلى معلومات. وبينما كان الصحافيون والمحررون هم الموجهين التقليديين

لجمع المعلومات ونشرها، فإن بيئة المعلومات المسطحة في 2012 تقوم الآن بنشر الأخبار العاجلة على الموقع الإلكتروني أولاً، وليس في مكتب الأخبار.

بل، وفي واقع الأمر، تزداد قوة العلاقة بين البيانات والصحافة حول العالم. ففي عصر البيانات الهائلة، تكمن الأهمية المتنامية لصحافة البيانات في قدرة ممارسيها على تقديم السياق والوضوح، وربما الأهم من ذلك إيجاد الحقيقة في الكم المتزايد من المحتوى الرقمي عالميا. وهذا لا يعني أن مؤسسات الإعلام الشامل اليوم لا تؤدي دوراً أساسياً. بل على العكس تماماً، ففي عصر المعلومات، يحتاج الصحافيون أكثر من أي وقت مضى إلى حفظ البيانات والتحقق منها وتصحيحها. وفي هذا السياق، فإن صحافة البيانات تصبح ذات أهمية كبرى بالنسبة للمجتمع.

واليوم، سيكون فهم البيانات الهائلة – وخاصة تلك التي تفتقر إلى التنظيم – هدفاً رئيسياً لعلماء البيانات حول العالم، سواء أكانوا يعملون في غرف الأخبار أو "وول ستريت" أو "سيليكون فالي". وعلى وجه الخصوص، سيمكن تحقيق هذا الهدف بشكل كبير عبر مجموعة متنامية من الأدوات المشتركة، سواء أكانت تستخدم من قبل تقنيي الحكومة لتطبيق نظام الحكومة المفتوحة في شيكاغو، أم تقنيي نظام الرعاية الصحية، أم مطوري غرفة الأخبار.

—— أليكس هاورد، أورايلي ميديا

حياتنا مكونة من البيانات

صحافة البيانات الجيدة أمر يصعب تحقيقه، وذلك لأن الصحافة الجيدة عموما صعبة. فهي تعني التوصّل إلى كيفية الحصول على البيانات وفهمها وكيفية العثور على الخبر. أحياناً يكون الطريق مسدوداً، وفي أحيان أخرى لا يوجد خبر مثير. وفي نهاية المطاف، إذا كانت المسألة تتعلق فقط بضغط الزر الصحيح، فلن تعتبر تلك صحافة. وفي عالم تتكون فيه حياتنا من بيانات مستمرة تصبح تلك البيانات أساسية من أجل إنشاء مجتمع حر وعادل.

—— كريس تاجار، قاعدة بيانات المؤسسات المفتوحة OpenCorporates

وسيلة لتوفير الوقت

ليس لدى الصحافيين وقت لإضاعته في تدوين الأشياء بأيديهم والسعي إلى الحصول على بيانات من ملفات PDF، لذا فمن القيّم جداً تعلُّم القليل عن الكتابة بالترميز "الكود" أو معرفة أماكن البحث عمن يمكنه المساعدة.

كان هناك صحفي من صحيفة فولا دي ساو باولو Folha de Sao Paulo يعمل على الميزانية المحلية، واتصل بي ليشكرنا على وضع حسابات بلدية ساو باولو على الإنترنت (الأمر الذي استغرق يومي عمل فقط من مبرمج واحد!). وقال أنه كان يدونها بيديه على مدى الثلاثة أشهر الماضية محاولاً التوصل إلى الخبر. كما أنني أذكر أننا توصلنا إلى حل لمشكلة ملفات PDF لمؤسسة كونتاس أبيرتاس Contas Abertas، وهي مؤسسة لرصد أخبار البرلمان: 15 دقيقة و15 سطراً من الكود حلت مشكلة تستغرق شهوراً من العمل.

—— بيدرو ماركون، مبرمج في منظمة الشفافية الدولية

جزء أساسي من أدوات الصحفي

أعتقد أنه من المهم أن نسلط الضوء على جانب 'الصحافة' أو إعداد التقارير فيما يتعلق 'بصحافة البيانات'. فلا يجوز أن يتعلق الأمر فقط بتحليل البيانات أو توضيحها مرئياً، بل ينبغي استخدامها كأداة للتقرب أكثر من حقيقة ما يحدث في العالم. أرى أن القدرة على تحليل البيانات وتفسيرها تعتبر أداة أساسيةً من أدوات الصحافة المعاصرة، لا فرعاً منفصلاً. وفي النهاية، يتعلق الأمر بالصحافة الجيدة وسرد الأخبار بأنسب طريقة.

تعتبر صحافة البيانات طريقة أخرى للتدقيق في شؤون العالم والتمتع بصلاحيات المساءلة. ومع توفر كم متزايد من البيانات، من المهم – الآن أكثر من أي وقت مضى – أن يدرك الصحافيون أساليب صحافة البيانات، التي يجب أن تكون إحدى أدوات أي صحفي: سواء أكان هذا من خلال تعلم استخدام البيانات بشكل مباشر، أم من خلال التعاون مع شخص يمكنه استخدامها.

تكمن القوة الحقيقية للبيانات في أنها تساعدك في الحصول على المعلومات التي قد يصعب العثور عليها أو إثباتها. والمثال الجيد على ذلك الخبر الذي أعدّه

ستيف دويغ والذي حلل فيه أنماط الخسائر الناتجة عن إعصار أندرو. لقد جمع طائفتين مختلفتين من البيانات: إحداهما ترسم خريطة لمدى الدمار الذي سببه الإعصار والأخرى توضح سرعات الرياح، مما أتاح له تحديد مناطق المباني الضعيفة وأعمال البناء الرديئة التي ساهمت في زيادة تأثير الكارثة. وقد حاز على جائزة بوليتزر عام 1993 (http://www.pulitzer.org/awards/1993) وهذا مثال رائع لما يمكن تحقيقه.

في أفضل الحالات، يمكنك استخدام البيانات لتحديد القيم الشاذة أو المجالات الهامة أو الأمور المفاجئة. ومن هذا المنطلق يمكن للبيانات أن تكون بمثابة خيط أو معلومة. وعلى الرغم من أن الأرقام قد تثير الاهتمام، إلا أن كتابتك عن البيانات لا تكفي، فلا تزال بحاجة إلى إعداد التقرير لشرح معناها.

— سينثيا أومورتشو، صحيفة فاينانشال تايمز

التكيف مع المتغيرات فيما يحيط بنا من معلومات

تجلب لنا التقنيات الرقمية الحديثة طرقاً جديدة لإنتاج المعرفة ونشرها في المجتمع. ويمكن فهم صحافة البيانات على أنها محاولة من الإعلام للتكيف والاستجابة للتغيرات فيما يحيط بنا من معلومات، بما في ذلك سرد الأخبار بقدر أكبر من التفاعل وتعدد الأبعاد، مما يتيح للقراء اكتشاف المصادر التي تعتمد عليها الأخبار، كما يشجعهم ذلك على المشاركة في عملية صنع وتقييم الخبر.

— سيزار فيانا، جامعة جوياس

طريقة لرؤية الأمور التي قد لا تراها في غياب البيانات

هناك بعض الأخبار التي يمكن فهمها وشرحها فقط عن طريق تحليل البيانات وتوضيحها بصرياً في بعض الأحيان. قد تظل العلاقات بين الأشخاص أو الكيانات المتعمتة بالسلطة غير معلنة، وقد تظل الوفيات الناتجة عن سياسات المخدرات مخفية، كما قد تتواصل السياسات البيئية التي تضر الطبيعة، ولكن كل ما سبق ذكره قد تغير بسبب البيانات التي حصل عليها الصحافيون وحللوها وقدموها للقراء. يمكن للبيانات أن تكون بسيطة مثل جدول بسيط أو سجل لمكالمات هاتفية، وقد

تكون معقدة مثل نتائج اختبارات المدارس أو بيانات العدوى في المستشفيات، ولكن في داخلها كلها أخبارا تستحق أن تُروى.

—— شيريل فيليبس، صحيفة سياتل تايمز

طريقة لسرد أخبار أكثر ثراءً

يمكننا رسم صور لحياتنا بأكملها باستخدام آثارنا الرقمية. بدءاً بما نستهلكه ونتصفحه وأماكن سفرنا ومواعيده والموسيقى المفضلة لدينا وحبنا الأول وإنجازات أطفالنا، وصولاً إلى أبعد أمانينا. يمكن تتبع كل هذا وجعله رقمياً وتخزينه في الشبكات السحابية ونشره. هذا الكون المكون من البيانات يمكن استغلاله لسرد الأخبار والإجابة عن الأسئلة وتناقل مفاهيم الحياة بطرق تتجاوز الآن إعادة تشكيل حتى أكثر الحكايات قوة ودقةً.

—— سارا سلوبين، صحيفة وول ستريت جورنال

لا تحتاج إلى بيانات جديدة لتحصل على سبق صحفي

أحياناً تكون البيانات عامة ومتاحة بالفعل، ولكن لم ينظر فيها أحد عن قرب. في حالة تقرير وكالة أسوشيتد برس المنشور في 4,500 صفحة من الوثائق التي رفعت عنها السرية والتي تصف ممارسات الشركات الأمنية الخاصة أثناء حرب العراق، حصل على المادة صحفي مستقل بتقديم طلبات إلى وزارة الخارجية الأمريكية بالحصول على المعلومات بموجب قانون 'حرية المعلومات'. وبمسح النتائج المكتوبة على ورق وتحميلها على الشبكة السحابية DocumentCloud، أتيح لنا القيام بتحليلنا الشامل.

—— جوناثان ستراي، مشروعOverview

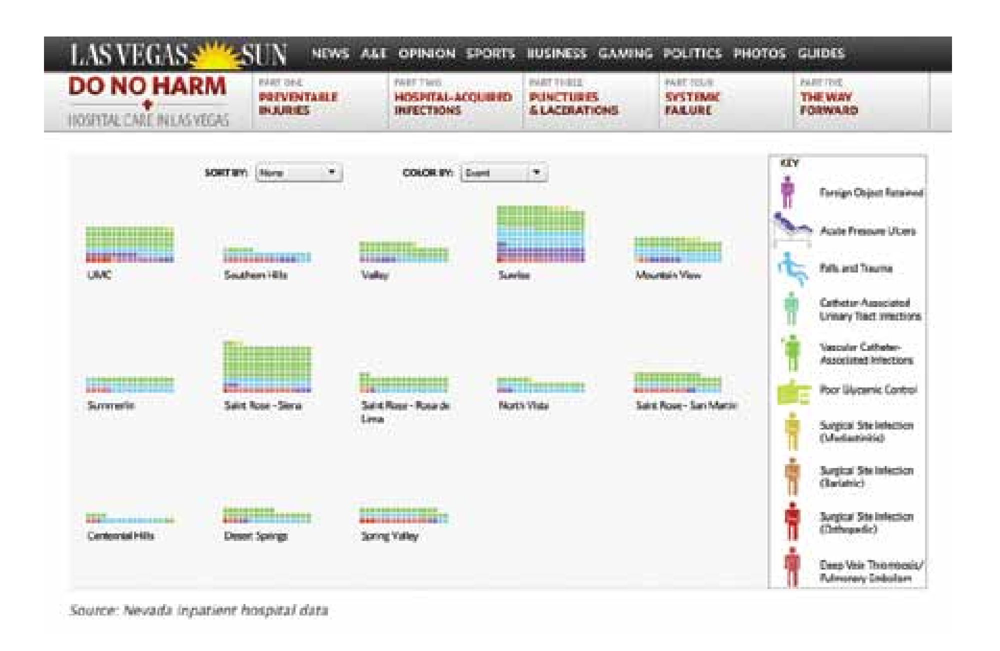
بعض الأمثلة المفضلة

طلبنا من بعض المساهمين أن يخبرونا عن أمثلتهم المفضلة بشأن صحافة البيانات وما أعجبهم فيها. ها هي الأمثلة:

سلسلة Do No Harm في صحيفة لاس فيغاس صن

مثالي المفضل هو سلسلة Do No Harm (لا تؤذِ) في صحيفة لاس فيغاس صن لعام 2010 حول الرعاية الصحية بالمستشفيات (http://www.lasvegassun.com/hospital-care/). قامت الصحيفة بتحليل ما يزيد على 2.9 مليون سجل لفواتير المستشفيات، مما كشف عن أكثر من 3600 حالة من الإصابات والعدوى والأخطاء الجراحية التي كان يمكن الوقاية منها. وقد حصلوا على البيانات عن طريق طلبات للسجلات الحكومية وحددوا أكثر من 300 حالة مات فيها المريض بسبب أخطاء كان يمكن تفاديها. تحتوي السلسلة على عناصر مختلفة، منها رسوم بيانية تفاعلية (http://bit.ly/lvsun-surgery) تسمح للقراء بالتصفح وفقاً للمستشفى الذي تزيد فيها الإصابات عن المعدل المتوقع، وخريطة زمنية (http://bit.ly/lvsun-infections) توضح انتشار العدوى في كل مستشفى على حِدة، بالإضافة إلى رسوم بيانية تفاعلية (http://bit.ly/lvsun-events) تتيح للمستخدمين تصنيف البيانات حسب الإصابات التي يمكن الوقاية منها أو حسب المستشفى لمعرفة أي المستشفيات يسبب ضرراً للناس. تعجبني السلسلة لأنها سهلة الفهم والتصفح. تمكن المستخدمين من استكشاف البيانات بمنتهى السهولة، كما أن لها تأثيراً حقيقياً: فقد رد المجلس التشريعي في ولاية نيفادا بإصدار ستة قوانين (http://bit.ly/lvsun-milestone). وعمل الصحافيون الذين شاركوا فيها جاهدين للحصول على البيانات وترتيبها. قام أحد هؤلاء الصحافيين، وهو أليكس ريتشاردز، بإعادة البيانات إلى المستشفيات وإلى مسؤولي الولاية (http://bit.ly/poynter-webgold) عشرات المرات لتصحيح الأخطاء.

— أنجليكا بيرالتا راموس، صحيفة لا ناسيون، (الأرجنتين)



نموذج 1–3. سلسلة **Do No Harm** في صحيفة لاس فيغاس صن

قاعدة بيانات رواتب موظفي الحكومة

يعجبني كثيراً العمل الذي تقوم به المؤسسات المستقلة الصغيرة يومياً، مثل بروبابليكا ProPublica أو صحيفة تكساس تريبيون، فلديهما صحفي بيانات عظيم اسمه ريان ميرفي. إذا كان لا بد لي أن أختار، سيقع اختياري على مشروع قاعدة بيانات رواتب موظفي الحكومة في صحيفة تكساس تريبيون (http://bit.ly/texastrib-employee). يجمع هذا المشروع 660,000 من رواتب الموظفين في قاعدة بيانات واحدة ليتمكن المستخدمون من البحث والمساعدة في توليد الأخبار. يمكنك البحث حسب الوكالة أو الاسم أو الراتب. المشروع بسيط وهادف ويتيح للجمهور المعلومات التي لا يمكن الوصول إليها. كما أنه سهل الاستخدام ويولد الأخبار تلقائياً. وهذا مثال رائع يوضح سبب حصول موقع الصحيفة على عدد كبير من الزيارات.

—— سايمون روجرز، صحيفة الغارديان



نموذج 1–4. رواتب موظفي الحكومة (تكساس تريبيون)

توضيح النص الكامل لسجلات حرب العراق مرئياً، وكالة أسوشيتد برِس

يعد عمل جوناثانستراي وجوليان برجس على سجلات حرب العراق (http://bit.ly/jstray-warlogs) بمثابة فتح مُلهم لتحليل النصوص والتوضيح المرئي باستخدام أساليب تجريبية، وهذا من أجل التعرف على المواضيع التي تستحق التعمق فيها من خلال مجموعة بيانات نصية ضخمة.

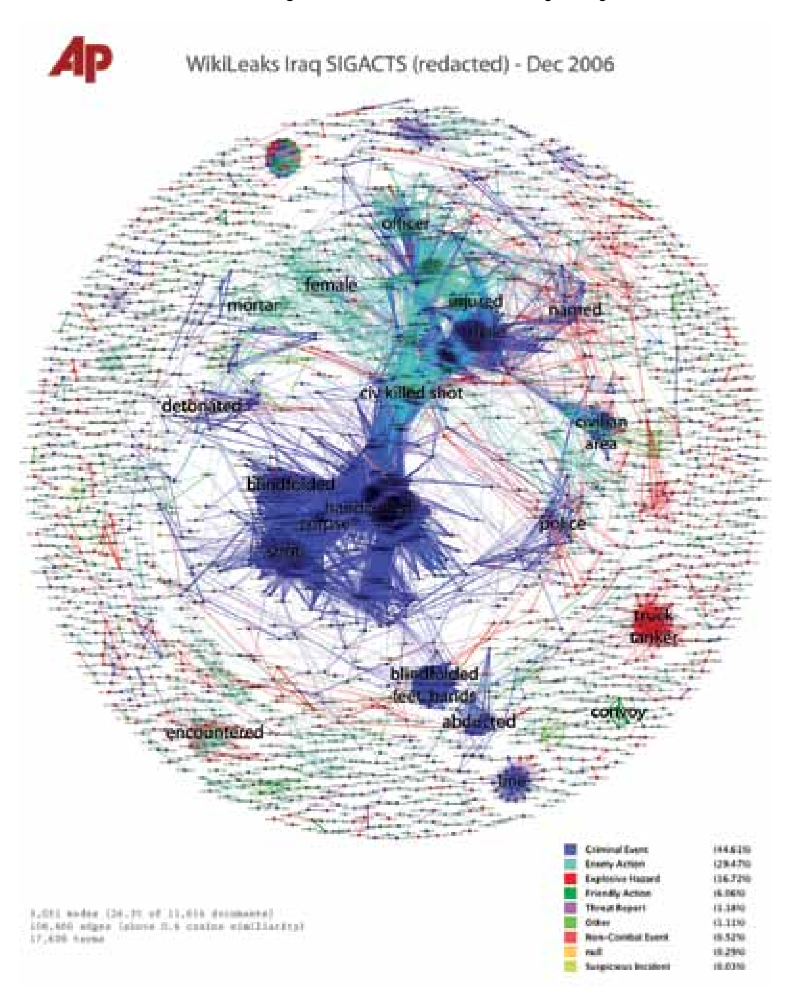
باستخدام أساليب ومناهج تحليلات النص، طور جوناثان وجوليان طريقة تظهر مجموعات من الكلمات البحثية الموجودة في آلاف من تقارير الحكومة الأمريكية حول حرب العراق والتي سربها موقع ويكيليكس في صورة مرئية.

وعلى الرغم من أن هناك قيوداً على الطريقة المطروحة وأن المنهج تجريبي، فقد طرحا منهجاً مبتكراً. وبدلاً من محاولة قراءة جميع الملفات أو مراجعة سجلات الحرب، فإنه – ومع وجود فكرة مسبقة حول ما يمكن أن يوجد عبر

إدخال كلمات بحثية معينة ومراجعة الناتج – يمكن باتباع هذا الأسلوب الوصول إلى ما هو مهم وتوضيحه بصرياً.

مع تزايد كم البيانات – النصية (كالبريد الإلكتروني والتقارير، إلخ) والرقمية – للجمهور، تزداد أهمية العثور على طرق لتحديد أهم المواضيع – وهذا مجال مثير ومتفرع من صحافة البيانات.

—— سينثيا أومورتشو، صحيفة فاينانشال تايمز

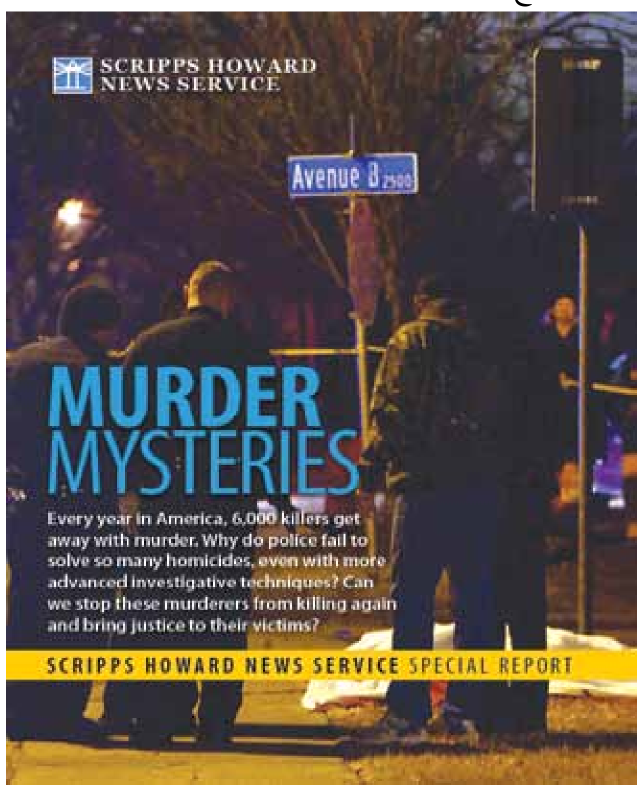


نموذج 1–5. تحليل سجلات الحرب (وكالة أسوشيتد برس)

جرائم القتل الغامضة

من أفضل مشاريع صحافة البيانات بالنسبة لي مشروع ‘جرائم القتل الغامضة’ لتوم هارجروف من سكريبس هاورد للخدمات الإخبارية Scripps Howard News Service (http://bit.ly/murder-mysteries). فباستخدام بيانات الحكومة وطلبات السجلات الحكومية، قام ببناء قاعدة بيانات مفصلة ديمغرافياً ومكونة مما يزيد عن 185,000 جريمة قتل لم تُحل بعد، ثم صمم نظاماً للبحث عن الأنماط مقترحاً احتمالات وجود سفاحين. يحتوي هذا المشروع على كل شيء: العمل الشاق، وقاعدة بيانات أفضل من تلك التي لدى الحكومة، وتحليل ذكي باستخدام أساليب العلوم الإنسانية، وعرض تفاعلي للبيانات على الإنترنت يتيح للقراء الاستكشاف بأنفسهم.

— ستيف دويغ، كلية والتر كرونكايت للصحافة، جامعة أريزونا

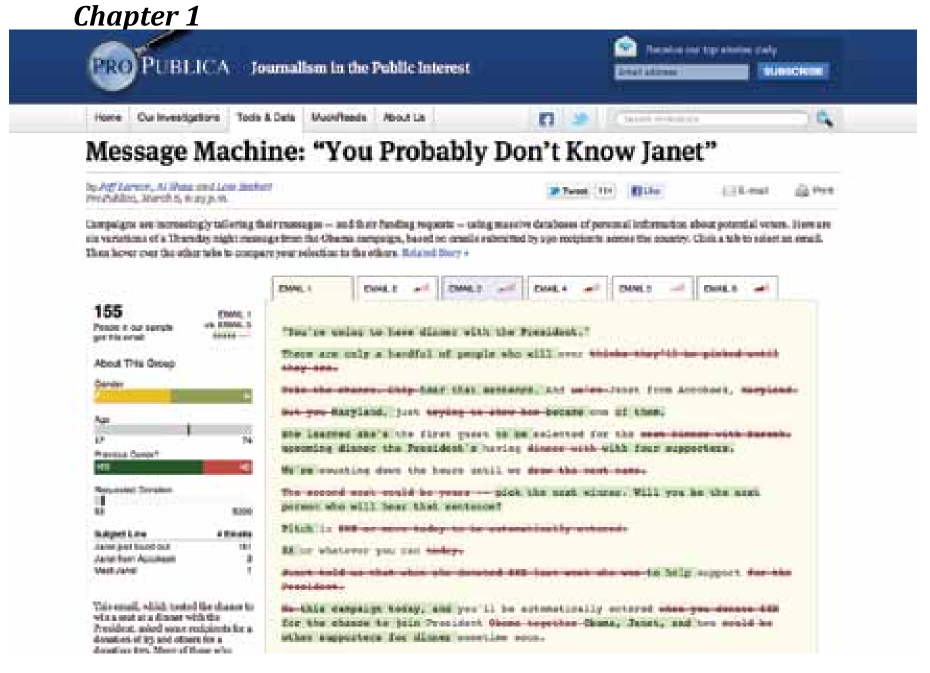


نموذج 1–6. جرائم القتل الغامضة (سكريبس هاورد للخدمات الإخبارية)

الرسائل الآلية

أعجبني كثيراً خبر الرسائل الآلية على موقع بروبابليكا (http://bit.ly/message-machine) ومدونة نيرد Nerd Blog Post (http://bit.ly/nerd-blog-post). بدأ الأمر حين أُثير فضول بعض مستخدمي تويتر بسبب تلقيهم رسائل إلكترونية مختلفة من حملة أوباما الانتخابية. وقد لاحظ العاملون في بروبابليكا هذا الأمر وطلبوا من جمهورهم إعادة إرسال أي رسائل يتلقونها من الحملة إليهم. كانت طريقة العرض أنيقة، وكان الفرق بسيطا بين عدة رسائل مختلفة تم إرسالها في ذلك اليوم. كان الأمر رائعاً لأنهم جمعوا بياناتهم بأنفسهم (كانت العينة صغيرة، إلا أنها كانت كبيرة بما يكفي لسرد الخبر). ولكن ما هو أروع هو أنهم سردوا قصة ظاهرة جديدة: ثمة كم كبير من البيانات يُستخدم في الحملات الانتخابية لاستهداف أفراد معينين بالرسائل. وهذه مجرد عينة مما سيأتي لاحقا.

—— بريان بوير، صحيفة شيكاغو تريبيون

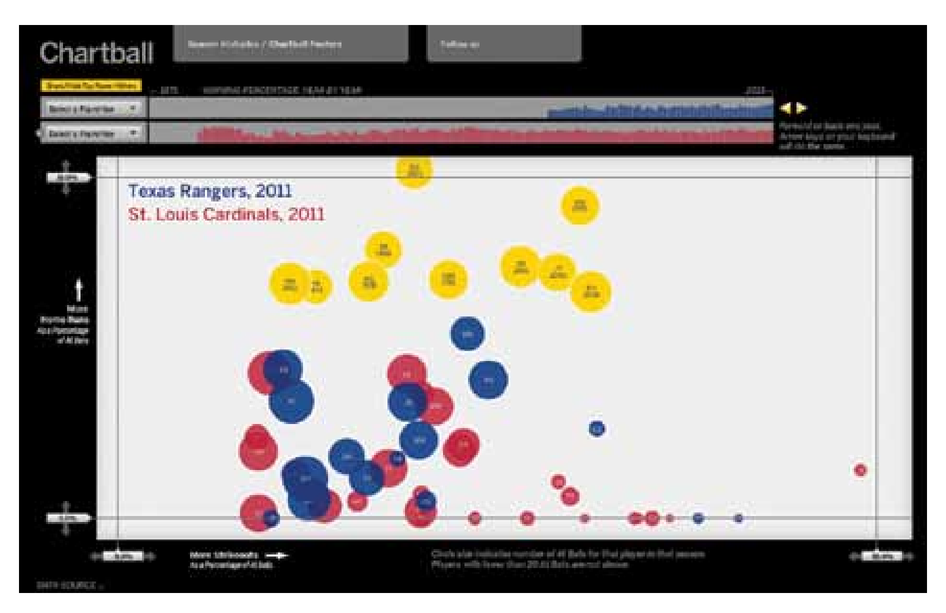


نموذج 1–7. الرسائل الآلية (بروبابليكا)

موقع Chartball

من أفضل مشاريع صحافة البيانات بالنسبة لي عمل أندرو غارسيا فيليبس على موقع Chartball ‏(http://www.chartball.com). أندرو من أكبر مشجعي الرياضة ولديه نهم للبيانات ورؤية رائعة للتصميم والقدرة على كتابة الكود. وفي هذا الموقع، لا يوضح التاريخ مرئياً فحسب، بل ينشر أيضاً تفاصيل نجاحات وإخفاقات اللاعبين والفرق. كما يضع سياقاً ورسوماً بيانية جذابة، بالإضافة إلى أن عمله عميق وممتع ومثير للاهتمام - هذا مع أنني لا أحب الرياضة كثيراً!

—— سارا سلوبان، صحيفة وول ستريت جورنال



نموذج 1–8. رسم بياني للنصر والهزيمة في الرياضة (موقع **Chartball**)

صحافة البيانات من منظور واسع

في أغسطس عام 2010 قمت وبعض الزملاء بتنظيم أحد أوائل المؤتمرات الدولية لصحافة البيانات (http://bit.ly/ddj-conf)، وقد عُقد في أمستردام. في هذا الوقت لم يكن هناك الكثير من النقاش حول هذا الموضوع، وكانت هناك

بضع مؤسسات فقط معروفة على نطاق واسع بعملها في هذا المجال.

كانت الطريقة التي تعاملت بها مؤسسات إعلامية مثل الغارديان ونيويورك تايمز مع الكم الهائل من البيانات الذي نشرته ويكيليكس من أهم الخطوات التي أبرزت مصطلح "صحافة البيانات". وفي ذلك الوقت، بدأ المصطلح يُستعمل بشكل أوسع (إلى جانب مصطلح 'الصحافة بمساعدة الكمبيوتر') ليصف كيفية استخدام الصحافيين للبيانات لتحسين تغطيتهم الصحفية ولتحفيز التقصي المتعمق في مواضيع معينة.

وبعد التحدث إلى صحافيي البيانات من ذوي الخبرة وأساتذة الصحافة على موقع تويتر (http://bit.ly/smfrogers-status)، بدا أن بدايات ظهور ما نعرفه الآن باسم صحافة البيانات كانت عام 2006 على يد أدريان هولوفاتي، مؤسس موقع EveryBlock، وهي خدمة معلومات تتيح للمستخدمين معرفة ما يحدث في منطقتهم وشارعهم. وفي مقاله القصير ''التغيير الأساسي الذي تحتاجه الصحف'' (http://www.holovaty.com/writing/fundamental-change/)، يقول إن على الصحافيين أن ينشروا بيانات منظمة ومقروءة آليا، إلى جانب ''كتلة النص الكبيرة'' والتقليدية:

"فلنقل مثلا إن هناك صحيفة نشرت خبراً عن حريق محلي. سيكون من الجيد بل والمدهش أن نتمكن من قراءة الخبر على الهاتف المحمول. إنها التكنولوجيا! ولكن ما أود التمكن من فعله حقًا هو تصفح الحقائق الخام المجردة وراء هذا الخبر، كلاً على حِدة، مع العديد من المراجع، وأساس لمقارنة تفاصيل الحريق مع تفاصيل الحريق السابق: التاريخ والوقت والمكان والضحايا ورقم مخفر الإطفاء والمسافة من دائرة الإطفاء والأسماء وسنوات الخبرة لرجال الإطفاء في موقع الحريق والوقت الذي استغرقه رجال الإطفاء للوصول والحرائق التالية أيا كان وقت وقوعها."

ولكن ما الذي يجعل هذا مختلفا عن أشكال الصحافة الأخرى التي تستخدم قواعد البيانات أو الكمبيوتر؟ كيف – ولأي مدى – تختلف صحافة البيانات عن أشكال الصحافة الأخرى التي عرفناها حتى الآن؟

الصحافة بمساعدة الكمبيوتر والصحافة الدقيقة

هناك تاريخ طويل وراء استخدام البيانات لإعداد التقارير وتقديم معلومات منظمة (إن لم تكن مقروءة آليا) إلى الجمهور. وربما تكون 'الصحافة بمساعدة الكمبيوتر' أو 'CAR' هي الأقرب لما نطلق عليه الآن اسم صحافة البيانات. فقد كانت أول أسلوب منظم ومنهجي لاستخدام الكمبيوتر في جمع وتحليل البيانات من أجل تحسين الأخبار.

استخدمت CAR للمرة الأولى عام 1952 من قبل شبكة سي. بي. إس. للتنبؤ بنتائج الانتخابات الرئاسية. ومنذ الستينيات، سعى الصحافيون المحققون (العاملون في الولايات المتحدة في أغلبهم) إلى مراقبة السلطة بشكل مستقل من خلال تحليل قواعد بيانات السجلات الحكومية بطرق علمية، وهو ما عُرف أيضاً باسم 'صحافة الخدمة العامة'. وقد سعى مؤيدو هذه الأساليب بمساعدة الكمبيوتر إلى إظهار التوجهات العامة وتعميم المعرفة الشعبية وكشف ظلم السلطات العامة والمؤسسات الخاصة. فعلى سبيل المثال، حاول فيليب ماير كشف ما كُتب عن أحداث الشغب عام 1967 في ديترويت كي يوضح أن المشاركة لم تقتصر على الجنوبيين الأقل تعليماً. وفي الثمانينيات كشفت سلسلة مقالات بيل ديدمان المعنونة 'لون المال' (The Color of Money) عن عنصرية ممنهجة في سياسات الإقراض في المؤسسات المالية الكبرى. وفي تحقيق بعنوان 'أين وقع الخطأ؟' سعى ستيف دويغ إلى تحليل أنماط الخسائر الناتجة عن إعصار أندرو في بداية التسعينيات من أجل فهم تأثير السياسات والممارسات المعيبة للتنمية العمرانية. لقد جلبت التقارير المبنية على البيانات خدمة عامة قيمة، وحققت لكثير من الصحافيين الفوز بجوائز شهيرة.

في أوائل السبعينيات، تم وضع مصطلح 'الصحافة الدقيقة' لوصف هذا النوع من جمع الأخبار: 'تطبيق مناهج البحث في العلوم الاجتماعية والسلوكية لممارسة العمل الصحفي' (من كتاب الصحافة الدقيقة الجديدة لفيليب ماير: http://bit.ly/precision-journalism). كان من المتصوّر أن تُمارس الصحافة الدقيقة في

مؤسسات الإعلام الرئيسية من قبل متخصصين مدربين في الصحافة والعلوم الاجتماعية. وقد ظهرت كرد على ’الصحافة الجديدة‘، وهو نوع من الصحافة طُبقت فيه بعض أساليب الأدب القصصي لإعداد التقارير. يقول ماير إن الأساليب العلمية لجمع البيانات وتحليلها – بدلاً من الأساليب الأدبية – هي ما تحتاجه الصحافة لإتمام بحثها عن الموضوعية والحقيقة.

يمكن فهم صحافة البيانات على أنها رد فعل على بعض أوجه القصور والضعف الشائعة في الصحافة: مثل الاعتماد على الإيجازات الصحفية التي تصوغها المؤسسات (التي وصفت فيما بعد بأنها "هراء الصحافة" أو Churnalism)، والانحياز إلى المصادر الرسمية، وما إلى ذلك مما يرى ماير أنه نابع من انعدام تطبيق أساليب المعلوماتية والطرق العلمية كاستطلاعات الرأي والسجلات الحكومية. وكما تم في الستينيات، استخدمت الصحافة الدقيقة لتمثيل المجموعات المهمشة وأخبارهم. يقول ماير:

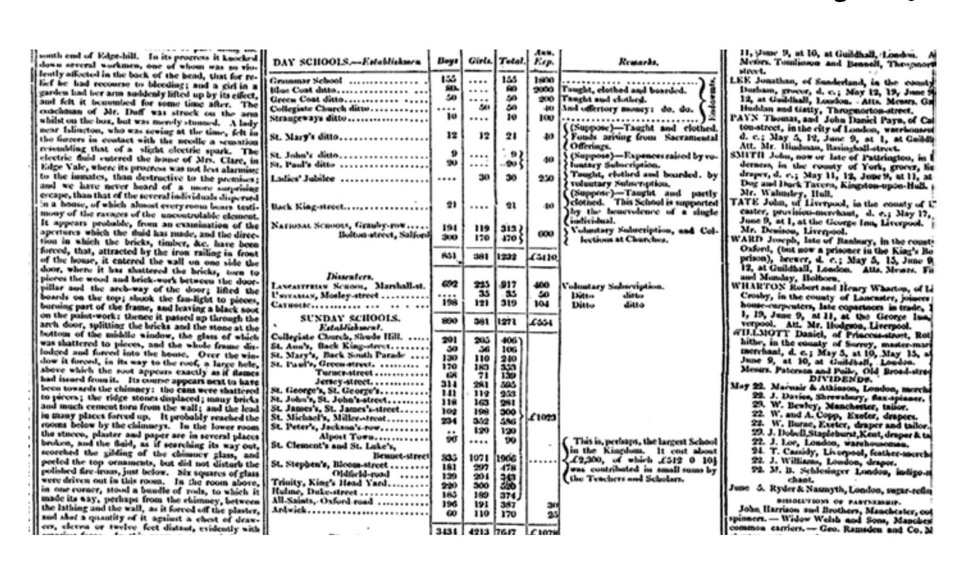
"كانت الصحافة الدقيقة وسيلة لتوسيع أدوات الصحفي ليتمكن من التدقيق الصحفي في المواضيع التي لم تكن متاحة أو كانت محدودة. كما كانت مفيدة خاصة في إعطاء صوت للأقليات والمنشقين الذين كانوا يناضلون من أجل حق التمثيل النيابي."

هناك مقال مؤثر (http://bit.ly/oxford-influential) نُشر في الثمانينيات حول العلاقة بين الصحافة والعلوم الاجتماعية يردد ما يدور في الحوار الحالي بشأن صحافة البيانات. يقول الكاتبان، وهما أستاذان في الصحافة في الولايات المتحدة، إن فهم الناس لماهية الأخبار أخذ في السبعينيات والثمانينيات ينتقل من المفهوم الأضيق عن ‘أحداث الأخبار’ إلى ‘تقارير الحالة’ (أو إعداد التقارير حول التوجهات الاجتماعية). وعلى سبيل المثال، يستطيع الصحافيون باستخدام قواعد بيانات تعداد السكان أو الدراسات الاستقصائية ”أن يتجاوزوا التقارير حول أحداث معينة ومنعزلة وأن يقدموا سياقا يعطي مغزى حقيقياً.»

كما هو متوقع، يعود استخدام البيانات لتحسين إعداد التقارير إلى بداية وجود البيانات. وكما أشار سايمون روجرز (http://bit.ly/facts-are-sacred)، فإن

أول أمثلة صحافة البيانات في صحيفة الغارديان يعود إلى عام 1821، وكان هذا جدولا مسربا عن المدارس في مانشستر يحتوي على قائمة بعدد الطلاب وتكاليف كل مدرسة. ووفقاً لروجرز، ساعد هذا الجدول للمرة الأولى على توضيح العدد الحقيقي للطلاب الذين يحصلون على التعليم المجاني، والذي كان أعلى كثيراً مما أظهرته الأرقام الرسمية.

ومن أوائل الأمثلة أيضاً في أوروبا تقرير فلورنس نايتينجايل الهام ’’وفيات الجيش البريطاني‘‘ (http://bit.ly/mortality-army) الذي نُشر عام 1858. ففي تقريرها للبرلمان، استخدمت الرسوم البيانية للدعوة إلى تحسين الخدمات الصحية في الجيش البريطاني. أشهر تلك الرسوم البيانية كان يسمى ’’coxcomb‘‘، وهو مجموعة من الأقسام يوضح كل منها المعدل الشهري للوفيات، مما أشار إلى أن أغلبية الوفيات كان سببها الأمراض التي كان يمكن الوقاية منها، وليس الرصاص.



south end of Edge-hill. In its progress it knocked down several workmen, one of whom was so violently affected in the back of the head, that for relief he had recourse to bleeding; and a girl in a garden had her arm suddenly lifted up by its effect, and felt it benumbed for some time after. The coachman of Mr. Duff was struck on the arm whilst on the box, but was merely stunned. A lady near Islington, who was sewing at the time, felt in the fingers in contact with the needle a sensation resembling that of a slight electric spark. The electric fluid entered the house of Mrs. Clare, in Edge Vale, where its progress was not less alarming to the inmates, than destructive to the premises; and we have never heard of a more surprising escape, than that of the several individuals dispersed in a house, of which almost every room bears testimony of the ravages of the uncontrolable element. It appears probable, from an examination of the apertures which the fluid has made, and the direction in which the bricks, timber, &c. have been forced, that, attracted by the iron railing in front of the house, it entered the wall on one side the door, where it has shattered the bricks, torn to pieces the wood and brick-work between the door-pillar and the arch-way of the door; lifted the boards on the top; shook the fan-light to pieces, burning part of the frame, and leaving a black soot on the paint-work: thence it passed up through the arch door, splitting the bricks and the stone at the bottom of the middle window, the glass of which was shattered to pieces, and the whole frame dislodged and forced into the house. Over the window it forced, in its way to the roof, a large hole, above which the soot appears exactly as if flames had issued from it. Its course appears next to have been towards the chimney: the cans were shattered to pieces; the ridge stones displaced; many bricks and much cement torn from the wall; and the lead in many places forced up. It probably reached the rooms below by the chimneys. In the lower room the stucco, plaster and paper are in several places broken, and the fluid, as if searching its way out, scorched the gilding of the chimney glass, and peeled the top ornaments, but did not disturb the polished fire-irons, just below. Six squares of glass were driven out in this room. In the room above, in one corner, stood a bundle of rods, to which it made its way, perhaps from the chimney, between the lathing and the wall, as it forced off the plaster, and shot a quantity of it against a chest of drawers, eleven or twelve feet distant, evidently with

DAY SCHOOLS.—*Establishment*	*Boys*	*Girls.*	*Total.*	*Ann. Exp.*	*Remarks.*
Grammar School	155	....	155	1800	.... (Endowments.)
Blue Coat ditto	80	....	80	2000	Taught, clothed and boarded. (Endowments.)
Green Coat ditto	50	....	50	200	Taught and clothed. (Endowments.)
Collegiate Church ditto	....	50	50	40	And offertory money: do. do. (Endowments.)
Strangeways ditto	10	....	10	100	.... (Endowments.)
St. Mary's ditto	12	12	24	40	(Suppose)—Taught and clothed. Funds arising from Sacramental Offerings.
St. John's ditto	9	....	9	40	(Suppose)—Expences raised by voluntary Subscription.
St. Paul's ditto	20	....	20		
Ladies' Jubilee	....	30	30	250	Taught, clothed and boarded. by voluntary Subscription.
Back King-street	21	....	21	40	(Suppose)—Taught and partly clothed. This School is supported by the benevolence of a single individual.
NATIONAL SCHOOLS, Granby-row	194	119	313	600	Voluntary Subscription, and Collections at Churches.
Bolton-street, Salford	300	170	470		
	851	381	1232	£5110	
Dissenters.					
LANCASTRIAN SCHOOL, Marshall-st.	692	225	917	400	Voluntary Subscription.
UNITARIAN, Mosley-street	....	35	35	50	Ditto ditto
CATHOLIC	198	121	319	104	Ditto ditto
	890	381	1271	£554	
SUNDAY SCHOOLS. *Establishment.*					
Collegiate Church, Shude Hill.	201	205	406	£1023	
St. Ann's, Back King-street	50	56	106		
St. Mary's, Back South Parade	130	110	240		
St. Paul's, Green-street.	170	183	353		
Turner-street	68	71	139		
Jersey-street.	314	281	595		
St. George's, St. George's	141	119	253		
St. John's, St. John's-street	118	163	281		
St. James's, St. James's-street	102	198	300		
St. Michael's, Miller-street	234	352	586		
St. Peter's, Jackson's-row	..	120	120		
Alport Town	90	....	90		
St. Clement's and St. Luke's, Bennet-street	835	1071	1906	......	This is, perhaps, the largest School in the Kingdom. It cost about £2,300, of which £512 0 10½ was contributed in small sums by the Teachers and Scholars.
St. Stephen's, Bloom-street	181	297	478		
Oldfield-road	139	204	343		
Trinity, King's Head Yard	220	300	520		
Hulme, Duke-street	185	189	374		
All-Saints, Oxford road	196	191	387	30	
Ardwick	60	110	170	25	
	3434	4213	7647	£1078	

11, June 9, at 10, at Guildhall, London. A
Messrs. Tomlinson and Bennett, Throgmort
street.
LEE Jonathan, of Sunderland, in the county
Durham, grocer, d. c.; May 12, 19, June 9
12, at Guildhall, London. Atts. Messrs. G
Hudson and Gatty, Throgmorton-street.
PAYN Thomas, and John Daniel Payn, of Ca
ton-street, in the city of London, warehouse
d. c.; May 5, 12, June 9, at 1, at Guildh
Att. Mr. Hindman, Basinghall-street.
SMITH John, now or late of Patrington, in H
derness, in the county of York, grocer, lin
draper, d. c.; May 11, 12, June 9, at 11, at
Dog and Duck Tavern, Kingston-upon-Hull.
Mr. Walmsley, Hull.
TATE John, of Liverpool, in the county of L
caster, provision-merchant, d. c.; May 17,
June 9, at 1, at the George Inn, Liverpool.
Mr. Denison, Liverpool.
WARD Joseph, late of Banbury, in the county
Oxford, (but now a prisoner in the King's Be
prison), brewer, d. c.; May 5, 15, June 9
12, at Guildhall, London. Atts. Messrs. Fi
and Munday, Holborn.
WHARTON Robert and Henry Wharton, of Li
Crosby, in the county of Lancaster, joiners
house-carpenters, late copartners in trade,
1, 19, June 9, at 11, at the George Inn,
verpool. Att. Mr. Hodgson, Liverpool.
WILLMOTT Daniel, of Princess-street, Rot
hithe, in the county of Surrey, master-mar
merchant, d. c.; May 5, at 10, May 15, a
June 9, at 10, at Guildhall, London.
Messrs. Paterson and Peile, Old Broad-stre

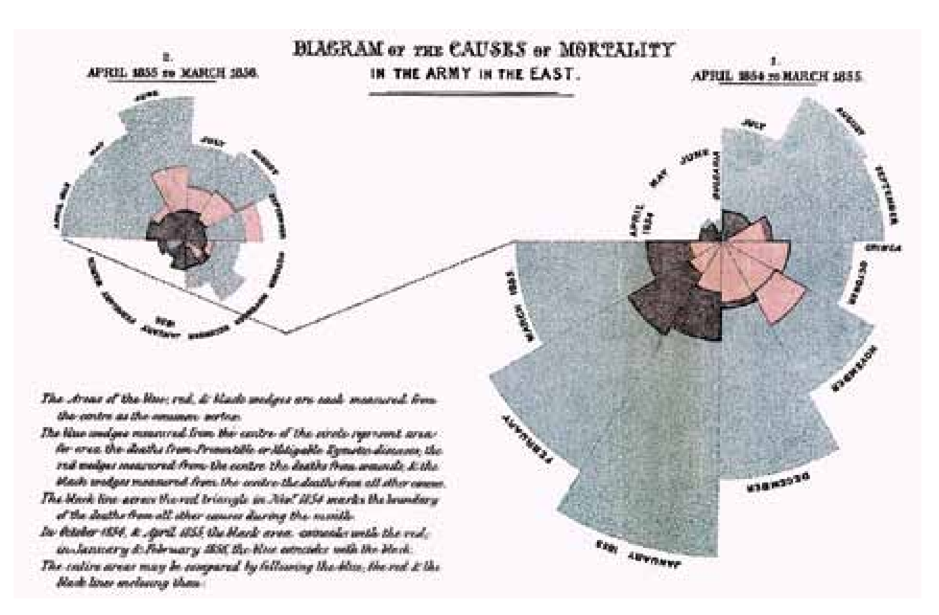
DIVIDENDS.

May 22. Marsair & Atkinson, London, merch
22. J. Davies, Shrewsbury, flax-spinner.
23. W. Bewley, Manchester, tailor.
22. W. and A. Copp, Exeter, drapers.
22. W. Burne, Exeter, draper and tailor.
22. J. Dobell, Staplehurst, Kent, draper & ta
22. J. Lee, London, warehouseman.
24. T. Cassidy, Liverpool, feather-merch
22. J. Williams, London, draper.
22. M. B. Schlesinger London, indigo-m
chant.
June 5. Ryder & Nasmyth, London, sugar-refin

DISSOLUTIONS OF PARTNERSHIP.

John Harrison and Brothers, Manchester, cot
spinners. — Widow Welsh and Sons, Manches
common carriers. — Geo. Ramsden and Co. M

نموذج 1–9. صحافة البيانات في صحيفة الغارديان عام 1821 (الغارديان)



نموذج 1–10. تقرير ’وفيات الجيش البريطاني‘ لفلورنس نايتينجايل (الصورة من ويكيبيديا)

صحافة البيانات والصحافة بمساعدة الكمبيوتر

يدور حالياً نقاش ’الاستمرارية والتغيير‘ حول مسمى ’صحافة البيانات‘ وعلاقته بتلك الممارسات الصحفية السابقة التي تستخدم تقنيات حاسوبية لتحليل مجموعات من البيانات.

يقول البعض إن هناك فرْقا بين CAR (الصحافة بمساعدة الكمبيوتر Computer Aided Reporting) وصحافة البيانات. ويقولون أن CAR أسلوب لجمع البيانات وتحليلها كوسيلة لتعزيز الصحافة (عادة ما تكون الصحافة الاستقصائية)، بينما تهتم صحافة البيانات بطريقة وضع البيانات في السياق الكامل للعمل الصحفي. من هذا المنطلق، تهتم صحافة البيانات بشكل أكبر بالبيانات نفسها بدلاً من استخدام البيانات كمجرد وسيلة لإيجاد الأخبار أو تعزيزها. لذا نجد أن مدونة بيانات الغارديان أو صحيفة تكساس تريبيون تنشران مجموعات البيانات إلى جانب الأخبار، أو حتى تنشران مجموعات البيانات وحدها ليحللها الناس ويستكشفونها.

هناك فرق آخر وهو أن الصحافيين الاستقصائيين كانوا في الماضي يعانون

من فقر في المعلومات المتعلقة بسؤال يحاولون الإجابة عنه أو بقضية يحاولون تناولها. وبينما لا يزال الحال كذلك بالطبع، هناك وفرة هائلة في المعلومات لا يعلم الصحافيون بالضرورة ماذا يفعلون بها. فهُم لا يعرفون كيفية الحصول على قيمة للبيانات التي بين أيديهم. ومثال أخير على ذلك هو قاعدة بيانات 'نظام المعلومات الموحدة على الإنترنت' (Combined Online Information System)، أكبر قاعدة بيانات لمعلومات الإنفاق في المملكة المتحدة. وقد سعى إليه طويلا دعاة الشفافية، ولكنه حير وأربك الكثيرين من الصحافيين عند إطلاقه. وكما كتب إليّ فيليب ماير مؤخراً: "حين كانت المعلومات نادرة، كرسنا معظم جهودنا للصيد والجمع. والآن مع وفرة المعلومات، فمعالجة البيانات أهم."

من ناحية أخرى، يقول البعض إنه لا يوجد فرق يذكر بين صحافة البيانات والصحافة بمساعدة الكمبيوتر. ومن المنطقي أن يكون هناك تاريخ حتى لدى أحدث ممارسات الإعلام، بالإضافة إلى ما هو جديد فيها. وبدلاً من النقاش حول ما إذا كانت صحافة البيانات جديدة كليا، هناك موقف أكثر إيجابية يتمثل في اعتبارها جزءاً من تقليد أقدم ولكنها تستجيب للظروف والأحوال الجديدة. وحتى إن لم يكن هناك فرق في الأهداف والأساليب، فإن ظهور مسمى 'صحافة البيانات' في بداية القرن يشير إلى بداية مرحلة جديدة، حيث الحجم الهائل من البيانات المتاحة على الإنترنت إلى جانب أدوات متطورة تركز على المستخدم بصفة شخصية وأدوات النشر الذاتي له والتشغيل الجماعي المنظم، مما يتيح للمزيد من الناس أن يعملوا باستخدام البيانات بشكل أسهل من أي وقت مضى.

صحافة البيانات تعني الإلمام الجماعي بالبيانات

تغير التقنيات الرقمية بشكل أساسي الطريقة التي تنشر بها المعلومات. وصحافة البيانات هي جزء واحد من نظام الأدوات والممارسات التي ظهرت مع مواقع البيانات وخدماتها. ويعد اقتباس المصادر وتعميمها من طبيعة الهيكل التشعبي لشبكة الإنترنت وهي الطريقة التي اعتدنا عليها لتصفح المعلومات اليوم. ومع الرجوع إلى أبعد من ذلك، فإن المبدأ الأساسي للهيكل التشعبي لشبكة الإنترنت هو مبدأ الاقتباس والاستشهاد بالمراجع المستخدم في الأعمال

الأكاديمية. فمن أهم الطرق التي يمكن للبيانات تحسين الصحافة بها اقتباس مصادر الخبر وبياناته وتعميمهما، وهذا ما يطلق عليه جوليان أسانج مؤسس موقع ويكيليكس اسم 'الصحافة العلمية.'

تمثل صحافة البيانات بشكل فعال عملية إرساء الديمقراطية الشاملة للمصادر والأدوات والأساليب والمناهج التي سبق أن استخدمها المتخصصون بشكل حصري، سواء أكانوا صحافيين استقصائيين أم علماء اجتماعيين أم إحصائيين أم محللين أم خبراء آخرين. وذلك من خلال تمكين الجميع من التنقيب بعمق في مصادر البيانات والعثور على المعلومات التي تهمهم، بالإضافة إلى التحقق من الادعاءات وتحدّي الافتراضات المعتادة. وفي حين أن الاقتباس والربط بمصادر البيانات حالياً أمر تختص به صحافة البيانات، نتجه نحو عالم تتداخل فيه البيانات بسلاسة مع نسيج الإعلام. ولدى صحافيي البيانات دور مهم في إزالة الحواجز أمام فهم البيانات والبحث فيها والتحقق منها، وزيادة إلمام القراء بالبيانات على نطاق واسع.

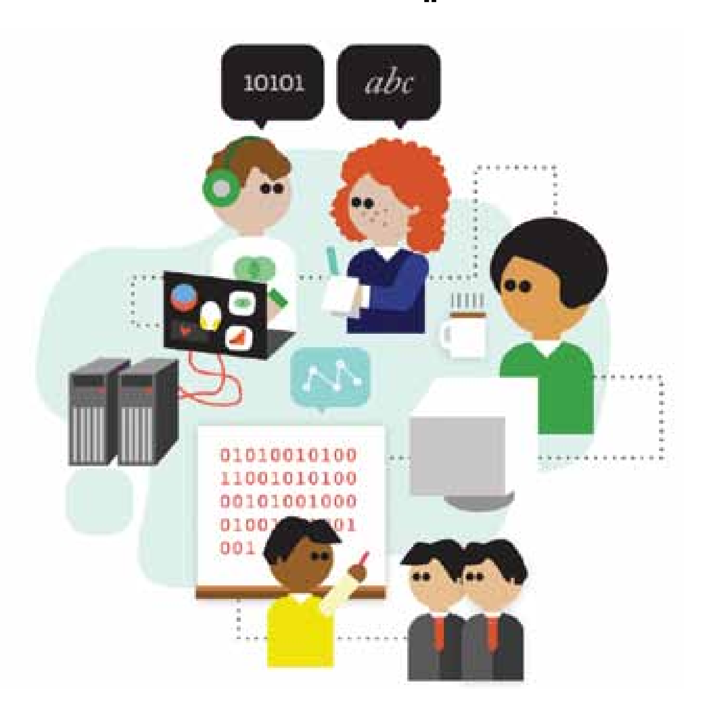
في هذه اللحظة يختلف كثيراً المجتمع الناشئ ممن يطلقون على أنفسهم اسم صحافيي البيانات عن مجتمع الصحافة بمساعدة الكمبيوتر CAR الأكثر نضوجاً. نأمل أن نرى في المستقبل روابط أكثر قوة بين هذين المجتمعين، مثلما نرى المنظمات غير الحكومية الجديدة وإعلام المواطنين مثل بروبابليكا ومكتب الصحافة الاستقصائية يعملون على التحقيقات الصحفية جنباً إلى جنب مع مؤسسات الأخبار التقليدية. وفي حين أن مجتمع صحافيي البيانات لديه طرق أكثر ابتكاراً لتقديم البيانات وعرض الأخبار، يظل المنهج النقدي التحليلي العميق لمجتمع CAR أمراً يمكن لصحافيي البيانات التعلم منه.

— ليليانا بونجيرو، مركز الصحافة الأوروبي

الفصل الثاني



في غرفة الأخبار



ما وضع صحافة البيانات داخل غرف الأخبار حول العالم؟ كيف استطاع كبار صحافيي البيانات إقناع زملائهم بأن نشر مجموعات البيانات، أو إطلاق تطبيقات إخبارية مبنية على البيانات، أفكار سديدة؟ هل ينبغي على الصحافيين تعلم كتابة الكود أو العمل جنباً إلى جنب مع المطوّرين الموهوبين؟ في هذا الباب سننظر في دور البيانات وصحافة البيانات في هيئة الإذاعة الأسترالية (ايه. بي. سي.) وشبكة بي. بي. سي. وصحيفتي شيكاغو تريبيون والغارديان إضافة إلى صحيفة دي تسايت الإلكترونية. ستتعلم كيفية العثور على المطوّرين الجيدين وتوظيفهم، وكيفية إشراك الناس في موضوع معين من خلال ملتقى المبرمجين (هاكاثون) والفعاليات الأخرى، كما ستتعلم كيفية التعاون عبر الحدود ونماذج الأعمال لصحافة البيانات.

صحافة البيانات في هيئة الإذاعة الأسترالية

هيئة الإذاعة الأسترالية (إيه. بي. سي.) تعدت السبعين عاما منذ تأسيسها، وهي الإذاعة الوطنية العامة في أستراليا، ويبلغ تمويلها السنوي حوالي مليار دولار أسترالي. تقدم الهيئة سبع شبكات راديو و60 محطة إذاعة محلية وثلاث قنوات تلفزيونية رقمية وقناة تلفزيونية دولية جديدة ومنصّة إلكترونية لتقديم هذا العرض المتزايد من المحتوى الرقمي والمحتوى الذي يتم إنشاؤه عن طريق المستخدم. حسب الإحصاءات الأخيرة، هناك ما يعادل 4500 موظف يعملون بنظام الدوام الكامل، منهم حوالي 70٪ يعملون لإنشاء المحتوى.

"نحن هيئة إذاعية وطنية فخورة باستقلالها، وعلى الرغم من أن الحكومة تمولنا، إلا أننا منفصلون عنها ولدينا صلاحيات مطلقة وفقاً للقانون. لدينا تقاليد تتمثل في صحافة الخدمة العامة المستقلة، وتعتبر هيئة إيه. بي. سي. من أكثر مؤسسات الأخبار الموثوقة في البلاد.

نحن نمر بأوقات مثيرة، وتحت إدارة مارك سكوت (المدير التنفيذي السابق للصحيفة)، تميز منشئو المحتوى في هيئة إيه. بي. سي. بـ 'سرعة البديهة وخفة الحركة' تطبيقاً لشعار المؤسسة.

بالطبع يسهل قول ذلك بينما يصعب تحقيقه. ولكن طُرحت مؤخراً مبادرة لتشجيع ذلك وهي منافسة بين الموظفين تتيح لهم الحصول على المال من أجل تطوير مشاريع متعددة المنصات.

وهكذا تم طرح أول مشروع لصحافة البيانات في هيئة إيه. بي. سي.

في أوائل عام 2010، ذهبت إلى اجتماع لأواجه ثلاثة من كبار المسؤولين عن الأفكار لأعرض عليهم اقتراحي.

كنت أفكر في هذا الأمر مليا منذ فترة، ملتهماً بنهم صحافة البيانات المتاحة على مدونة الغارديان والتي أضحت الآن أسطورية، وكانت تلك مجرد البداية.

كانت حجتي هي أنه خلال خمس سنوات – بلا شك – سيكون لدى هيئة

ABC وحدة صحافة بيانات خاصة بها. كنت أرى أن الأمر حتمياً. ولكن كيف يتحقق ذلك، ومن الذي سيبدأ في تحقيقه؟

بالنسبة للقراء الذين لا يعرفون هيئة إيه. بي. سي.، تخيلوا بيروقراطية هائلة متراكمة منذ 70 عاماً. كانت الهيئة تقدم أساساً خدمة الراديو والتلفزيون. ومع ظهور الإنترنت في العقد الأخير، تحول المحتوى إلى نصوص وصور وقدر من التفاعل الذي لم يكن تخيله ممكناً من قبل. أجبر الفضاء الإليكتروني الهيئة على إعادة التفكير في توزيع المال ونوعية المحتوى.

وهذا بالطبع عمل في طور التنفيذ.

ولكن، كان هناك أمر آخر يحدث في مجال صحافة البيانات. بدأت الحكومة الإلكترونية Government 2.0 (والتي أدركنا أنهم يتجاهلونها عموما في أستراليا) في تقديم طرق جديدة لسرد الأخبار الدفينة في البيانات الرقمية.

قلت لهم كل هذا أثناء عرضي للاقتراح. وقلت أيضاً إننا نحتاج إلى تحديد مجموعة مهارات جديدة وتدريب الصحافيين على استخدام أدوات جديدة. كنا نحتاج إلى مشروع لنبدأ العمل.

ثم منحوني المال.

في 24 نوفمبر 2011، أطلق مشروع إيه. بي. سي. المتعدد المنصات وموقع إيه. بي. سي. الإلكتروني للأخبار، مشروع "غاز الفحم بالأرقام" (CoalSeam Gas by the Numbers) (http://bit.ly/abc-coal).

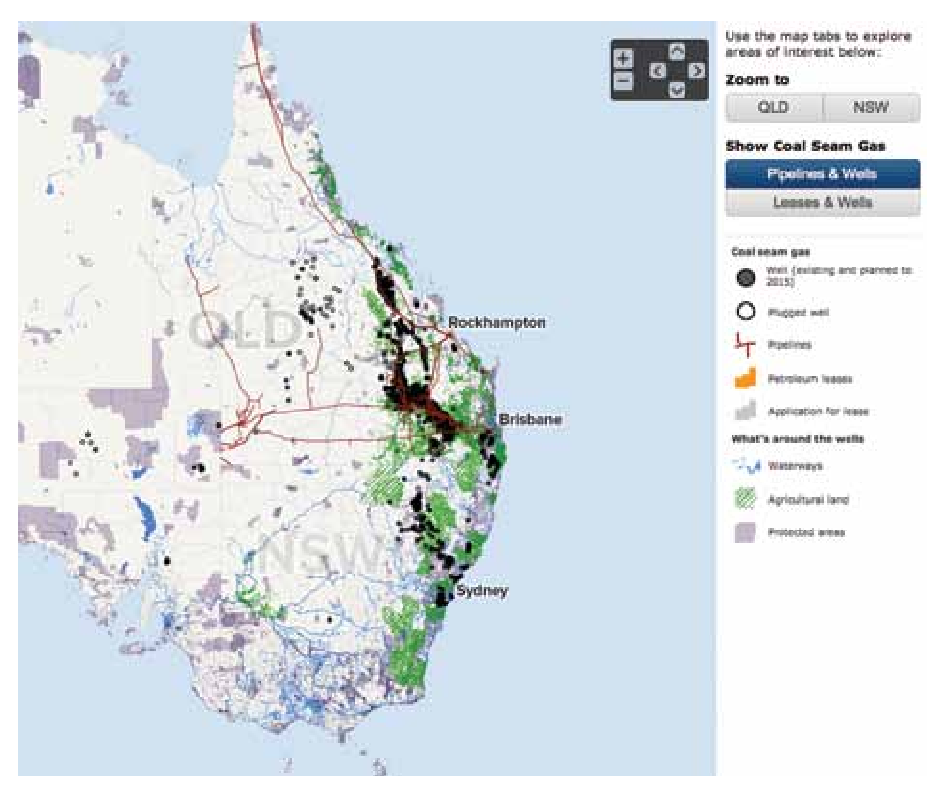


نموذج 2–1. 'غاز الفحم بالأرقام' (موقع إيه. بي. سي. الإلكتروني للأخبار)

كان مكوناً من خمس صفحات من الخرائط التفاعلية والإيضاحات المرئية للبيانات والنصوص.

ولم يكن عبارة عن صحافة بيانات فقط، بل كان مزيجاً من أنواع الصحافة التي نشأت عبر تفاعل أعضاء الفريق مع هذا الخبر الذي أصبح أحد أهم القضايا في أستراليا.

كانت درة الموقع الخريطة التفاعلية التي توضح آبار غاز الفحم وعقود إيجارها، مما أتاح للمستخدمين البحث حسب الموقع والتنقل بين طرق العرض لرؤية عقود الإيجار أو الآبار. ومن خلال تكبير الصورة، تمكن المستخدمون من معرفة شركة التنقيب وحالة الآبار وتاريخ الحفر. كما أظهرت خريطة أخرى موقع نشاط غاز الفحم مقارنة بموقع أنظمة المياه الجوفية في أستراليا.



نموذج 2–2. خريطة تفاعلية توضح آبار الغاز وعقود إيجارها في أستراليا
(موقع إيه. بي. سي. الإلكتروني للأخبار)

كان لدينا توضيحات مرئية للبيانات تتناول على وجه الخصوص قضية نفايات الملح وإنتاج المياه، كنا سنطلقها اعتماداً على السيناريو الذي يحدث.

كما حقق جزء آخر من المشروع في إلقاء مواد كيميائية في نهر محلي.

فريقنا:

- مطوّر ومصمم مواقع (ويب)
- صحفي رئيسي
- باحث بدوام جزئي لديه خبرة في استخراج البيانات وجداول إكسل وتنقية البيانات
- صحفي ثانوي بدوام جزئي
- منتج تنفيذي/ مستشار
- مستشار أكاديمي لديه خبرة في التنقيب واستخراج البيانات والتوضيح المرئي لها ويتمتع بمهارات بحثية متقدمة
- خدمات من مدير مشروع ومساعدة إدارية من وحدة إيه. بي. سي. المتعددة المنصات
- ومما يجدر ذكره، كانت لدينا أيضاً مجموعة متمرسة من الصحافيين وأصحاب المهن الأخرى الذين قمنا باستشارتهم عند الضرورة.

من أين حصلنا على البيانات؟

استخرجنا بيانات الخرائط التفاعلية من ملفات ‘shapefiles’ (وهو نوع معروف من الملفات من أجل البيانات الجغرافية) وقمنا بتحميلها من المواقع الإلكترونية الحكومية.

وقد حصلنا على البيانات الأخرى التي تخص الملح والمياه من تقارير متنوعة.

أما البيانات التي تخص إطلاق المواد الكيميائية، فقد حصلنا عليها من التراخيص البيئية التي أصدرتها الحكومة.

ماذا تعلمنا؟

كان مشروع ’غاز الفحم بالأرقام‘ مشروعا طموحاً في محتواه ومقاييسه.

والأهم في ذهني هو السؤال حول ما تعلمناه، وكيف يمكننا تطبيقه بشكل مختلف المرة القادمة؟

لقد جمعت صحافة البيانات في غرفة واحدة الكثيرين ممن لا يلتقون عادة في "إيه. بي. سي": أي الصحافيين والمبرمجين. عديد منا لم يكن يتحدث لغة الآخر أو حتى يقدّر ما تقوم به المجموعة الأخرى، فصحافة البيانات تتسم بانها فوضوية وهدّامة!

الأمور العملية:

- الاشتراك في الموقع بين أفراد الفريق أمر في غاية الأهمية. فقد كان كل من المطوّر والمصمم خارج موقع العمل وكانا يأتيان لحضور الاجتماعات. وهذا بالتأكيد ليس أفضل حل! يجب وضعهم جميعا في نفس الغرفة مع الصحافيين.
- كان المنتج التنفيذي المستشار أيضاً في طابق آخر بالمبنى. كنا في حاجة لأن نكون أقرب، للحفاظ على عنصر التواجد الدائم.
- يجب اختيار خبر يعتمد فقط على البيانات.

الصورة الكاملة: بعض الأفكار

ينبغي على المؤسسات الإعلامية الكبرى أن تشارك في بناء القدرات للتغلب على تحديات صحافة البيانات. هناك في اعتقادي الكثير من خبراء الكمبيوتر والمبرمجين المختبئين في الأقسام التقنية لمؤسسات الإعلام ويتوقون للخروج منها. لذا نحتاج إلى ورشات يلتقي فيها الخبراء من أصحاب القدرات الكامنة وشباب الصحافيين ومطوّرو الويب والمصممون الموهوبون ليخرجوا ويعملوا مع الصحافيين الأكثر خبرة لتبادل المهارات والتوجيه. المهمة: حمّل مجموعة البيانات وأبدأ في العمل!

وبطبيعة الحال فإن صحافة البيانات متعددة التخصصات، وتتكون فرقها من أشخاص لم يكونوا ليعملوا معا فيما مضى. ولكن الفضاء الرقمي أزال الحواجز.

بما أننا نعيش في كيان سياسي ممزق ومشتت وتنعدم به الثقة، وأصبح فيه نموذج العمل الذي كان يقدم صحافة مستقلة ومهنية – على الرغم من عيوبه –

على وشك الانهيار، فلا بد أن نسأل أنفسنا – كما يفعل الكثيرون الآن – كيف سيبدو العالم بدون صحافة مجدية؟ وهنا يجدر ذكر ما قاله الصحافي والمفكر الأمريكي والتر ليبمان في العشرينيات إن "من المسلم به أنه لا يمكن وجود رأي عام سليم بدون الحصول على الأخبار." وهذه العبارة لا تزال صحيحة حتى الآن. ففي القرن الحادي والعشرين، يقضي الجميع وقتهم في عالم المدونات. ويصعب التفريق بين المخادعين والكاذبين والمنافقين وأصحاب المصالح وبين الصحافيين المحترفين. أي أنه يمكن جعل أي موقع أو مصدر تقريباً موثوقاً به ولامعاً وصادقاً في الظاهر. كما أن الأسماء الكبرى الموثوق بها تنزوي في الظلام حيث تصبح مهددة بالانقراض. وفي هذه المساحة الجديدة للصحافة غير المرغوب فيها، يمكن للروابط الرقمية على النت أن تواصل نقل القارئ إلى مصادر أخرى عديمة الفائدة ولكنها مبهرة وتستمر الروابط في العودة إلى تلك المتاهة الرقمية. إن المصطلح الفني لهذا هو: "هراء مراوغ يحير العقول".

قد يظن البعض انه في هذا الفضاء الرقمي يمكن لأي شخص أن يصبح مصدرا لسرد الأخبار، وهذا ليس صحيحاً. فإذا أمكن للصحافة الاحترافية أن تبقى – وأعني هؤلاء الذين يتبنون سرد الأخبار بشكل أخلاقي ومتوازن وجريء وباحث عن الحقيقة ومراع لها – يجب على هذه الحرفة أن تفرض نفسها من جديد في الفضاء الرقمي، لأن صحافة البيانات هي مجرد أداة أخرى تمكننا من تصفح الفضاء الرقمي، حيث نستطيع رسم الخرائط للخبر ونقله وتدقيقه وتصفيته ورؤيته وسط كل تلك الأصفار والآحاد. وفي المستقبل، سنعمل جنباً إلى جنب مع المبرمجين والمطوّرين والمصممين وكاتبي الشفرة (الكود). هذا التحوّل يتطلب بناء جاداً للقدرات. حيث نحتاج إلى مديري أخبار 'يفهمون' الصحافة الرقمية من أجل البدء في البناء.

—— ويندي كارلايل، هيئة الإذاعة الأسترالية

صحافة البيانات في شبكة بي. بي. سي

يمكن لمصطلح 'صحافة البيانات' أن يغطي مجموعة من التخصصات ويُستخدم بطرق مختلفة في مؤسسات الأخبار، لذا قد يكون من المفيد أن نحدد

ما نعني بـ ’صحافة البيانات‘ في شبكة بي. بي. سي. يغطي المصطلح بشكل عام المشاريع التي تستخدم البيانات لتقوم بأحد أو بعض الأمور التالية:

- تمكين المتلقي من اكتشاف المعلومات التي تهمه شخصياً.
- الكشف عن خبر استثنائي هام وغير معروف مسبقاً.
- مساعدة المتلقي على فهم قضية معقدة بشكل أفضل.

هذه التصنيفات قد تتداخل، وفي بيئة إلكترونية، قد تستفيد أحياناً من قدر ما من التوضيح المرئي.

اجعلها شخصية

على الموقع الإلكتروني لأخبار بي. بي. سي كنا ولا نزال نستخدم البيانات لتوفير الخدمات والأدوات لمستخدمينا منذ أكثر من عشر سنوات.

من أكثر الأمثلة ملاءمة، وهذا نشرناه للمرة الأولى عام 1999، جداول دوري المدارس (http://bbc.in/school–league–tables) التي تستخدم بيانات تنشرها الحكومة سنوياً. يمكن للقراء العثور على المدارس المحلية من خلال إدخال الرمز البريدي ومقارنتها وفقاً لمجموعة من المؤشرات. يعمل صحافيو التعليم أيضاً مع فريق التطوير لاستخراج البيانات للأخبار قبل النشر.

عندما بدأنا في ذلك لم يكن هناك موقع إلكتروني واحد يوفر للناس طريقة يتحققون بها من البيانات. ولكن بما أن وزارة التعليم لديها الآن خدمات المقارنة الخاصة بها، انتقل ما نقدمه إلى التركيز أكثر على الأخبار التي تأتي من البيانات.

يكمن تحدي هذا المجال في توفير الوصول إلى البيانات التي تتمتع باهتمام واضح من الناس. ومثال أخير على مشروع كشفنا فيه عن مجموعة بيانات ضخمة لا تكون متاحة للجمهور عادة: ذلك هو التقرير الخاص بـ ’كل الوفيات على كل الطرق‘ (http://bbc.in/road–deaths)، حيث قدمنا البحث عن طريق الرمز البريدي ليتمكن المستخدمون من العثور على مواقع جميع حوادث الوفيات على كل طرق المملكة المتحدة في العقد الأخير.

وقد وضعنا توضيحات مرئية للحقائق والأرقام الرئيسية (/http: //bbc.in

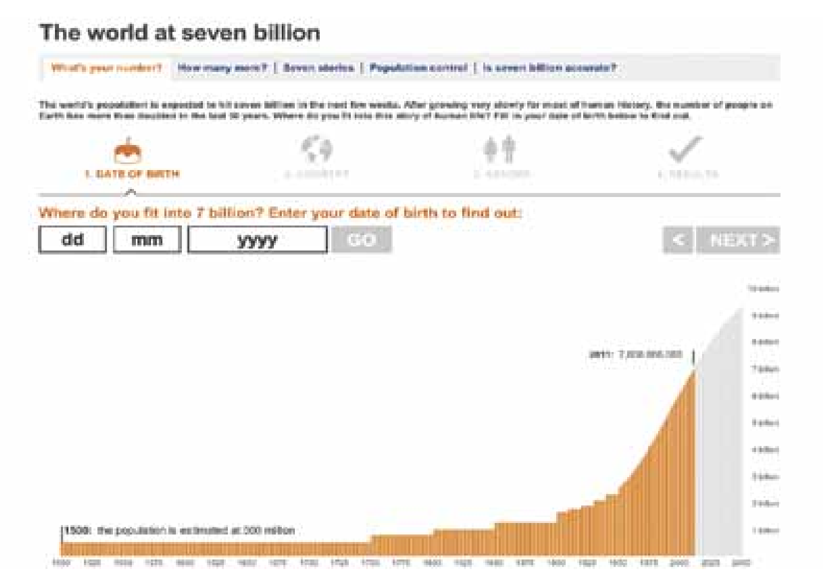
police-data) التي حصلنا عليها من بيانات الشرطة، ولإضفاء الحيوية والطابع الإنساني على المشروع، عملنا مع جمعية موظفي الإسعاف في لندن وراديو وتلفزيون بي. بي. سي لندن لتتبع حوادث الطرق في أنحاء العاصمة أثناء حدوثها. كان هذا يُنشر مباشرة على الإنترنت (http://bbc.in/road-deaths-feed)، وعلى موقع تويتر باستخدام الهاشتاج #crash24 وتم رسم خرائط للحوادث أثناء نشر أخبارها (http://bbc.in/road-deaths-map).

أدوات بسيطة

بالإضافة إلى توفير طرق لتصفح مجموعات البيانات الضخمة، نجحنا أيضاً في وضع أدوات بسيطة توفر للمستخدمين مقاطع من المعلومات التي تهمهم شخصياً. تحظى هذه الأدوات بإعجاب الذين يعانون من ضيق الوقت والذين لا يودون تصفح التحليلات الطويلة. وقد بدأنا في إدراج القدرة على نشر "حقائق شخصية" على أنها من الأمور الاعتيادية.

ومن الأمثلة الطريفة على هذا النهج تقريرنا الخاص بـ 'سبعة مليارات نسمة في العالم: ما هو ترتيبك؟' (http://bbc.in/KQsSzB)، والذي نُشر تزامناً مع التاريخ الرسمي الذي تعدّى فيه تعداد سكان العالم السبعة مليارات نسمة. وعبر إدخال تاريخ الميلاد، يتمكن المستخدم من معرفة 'ترتيبه' في تعداد سكان العالم عند ولادته، ثم المشاركة بنشر هذا الترتيب على مواقع تويتر أو فيسبوك. استخدم هذا التطبيق بيانات من صندوق الأمم المتحدة للسكان. كان التطبيق رائجاً وأصبح أكثر رابط تتم مشاركته على موقع فيسبوك في المملكة المتحدة عام 2011.

ومن الأمثلة الأخيرة أيضاً أداة حساب الميزانية من بي. بي. سي (http://bbc.in/JepssY) الذي مكّن المستخدمين من معرفة مدى تحسن حالتهم أو تدهورها حين تدخل ميزانية وزير المالية حيز التنفيذ، ثم يمكنهم المشاركة بنشر الرقم. وقد عملنا مع شركة المحاسبة KPMG LLP التي وفرت لنا حسابات مبنية على الميزانية السنوية، ثم عملنا بجد لوضع واجهة تفاعلية جذابة لتشجيع المستخدمين على إكمال المهمة.



نموذج 2–3. سبعة مليارات نسمة في العالم، وما ترتيبك أنت (بي. بي. سي.)

التنقيب في البيانات

ولكن أين الصحافة من كل ذلك؟ إن التعريف التقليدي لصحافة البيانات هو العثور على الأخبار وسط البيانات. هل هناك خبر حصري مدفون في قاعدة البيانات؟ هل الأرقام دقيقة؟ هل تثبت أو تدحض أي مسألة؟ كل هذه أسئلة يجب أن يطرحها صحفي البيانات، أو الصحفي الذي يستعين بكمبيوتر، على نفسه. ولكن يمكن إهدار الكثير من الوقت في فرز قاعدة بيانات هائلة والتدقيق فيها على أمل العثور على شيء مميز.

في هذا المجال، وجدنا أنه من المثمر الاشتراك مع فرق استقصائية أو برامج لديها الخبرة والوقت اللازمين لتقصّي الخبر. وقد قام برنامج 'بانوراما' المعني بأحداث الساعة في بي. بي. سي بقضاء عدة أشهر في العمل مع مركز الصحافة الاستقصائية لتجميع البيانات حول رواتب القطاع العام. وكانت النتيجة فيلماً وثائقياً تلفزيونياً وتقريراً خاصاً على الإنترنت اسمه 'رواتب القطاع العام: الأرقام' (http://bbc.in/IKPrL2)، حيث نُشرت جميع البيانات وتم توضيحها مرئياً مع تحليل لكل قطاع على حِدة.

بالإضافة إلى الشراكة مع الصحافيين الاستقصائيين، كان من الضروري الوصول إلى صحافيين مُلمين بالأرقام ولديهم معرفة متخصصة. عندما قام زميل في الفريق بتحليل بيانات تخفيض الميزانية التي نشرتها الحكومة، استنتج أنها كانت تجعلها تبدو أكبر مما هي عليه في الواقع. كانت النتيجة خبراً حصرياً تحت عنوان 'فهم البيانات' (http://bbc.in/LcuGFV) ومعه توضيح مرئي (http://bbc.in/IIADrj) وقد حصلا على جائزة الجمعية الملكية للإحصاء.

فهم المشاكل

ولكن ليس من الضروري أن تكون صحافة البيانات خبراً حصرياً لم يلحظه الآخرون. فعمل فريق التوضيح المرئي للبيانات يتكون من المزج بين تصميم رائع وسرد تحريري واضح لإعطاء مستخدم البيانات تجربة مذهلة. كما أن استخدام التوضيح المرئي للبيانات الملائمة يمكن أن يوفر فهماً أفضل لقضية أو خبر ما، وعادة ما نستخدم هذا النهج في سرد الأخبار في بي. بي. سي. أحد الأساليب التي استخدمناها في تتبع عدد الدعاوى القضائية في المملكة المتحدة تتبع خيوط البيانات الأكثر سخونة وتحديد معالمها عبر فترة زمنية ممتدة من الوقت لتقديم رؤية واضحة للتغيرات التي طرأت (http://bbc.in/KF7IKU).

يكشف تقرير بيانات 'شبكة ديون منطقة اليورو' (http://bbc.in/IIAH-HI) عن شبكة معقدة من القروض بين الدول، ويساعد في شرح قضية معقدة بطريقة مرئية وهذا باستخدام الألوان والأسهم النسبية مع نصوص واضحة. ومن المهم تشجيع القارئ على استكشاف التقرير أو متابعة القصة وألا يشعر أبداً بالارتباك بسبب الأرقام.

نظرة عامة على الفريق

يتكون فريق صحافة البيانات في موقع بي. بي. سي للأخبار من حوالي 20 صحفياً ومصمماً ومطوّراً.

بالإضافة إلى مشاريع البيانات وتوضيحها مرئياً، يقدم الفريق الرسوم البيانية

والتقارير التفاعلية المتعددة الوسائط على الموقع الإلكتروني للأخبار. كل هذه الأمور تجتمع لتكوّن أساليب لسرد الأخبار نطلق عليها 'الصحافة المرئية'. ليس لدينا من نطلق عليهم تحديداً 'صحافيي البيانات'، ولكن ينبغي على جميع أفراد الطاقم التحريري أن يجيدوا استخدام التطبيقات الأساسية لجداول البيانات مثل إكسل وأدوات جوجل لتحليل البيانات.

تعتبر القدرات الفنية والإرشادات التي يقدمها مطوّرونا ومهارات التوضيح المرئي لمصممينا جميعها أساسية للعمل على أي مشروع يرتبط بالبيانات. وبما أننا جميعا إما صحافيون أو مصممون أو مطوّرون في المقام الأول، فنحن نواصل العمل بجد لزيادة فهمنا وتحسين مهاراتنا في مجالات خبرة بعضنا البعض.

من أهم برامج التحقق من البيانات برنامج إكسل Excel ومحرر مستندات جوجل Google Docs وجداول فيوجن Fusion Tables، وبدرجة أقل، استخدم الفريق أيضاً برامج MySQL وقواعد بيانات Access وSolr للتحقق من قواعد البيانات الضخمة، بينما استخدم برامج RDF وSPARQL للبدء في البحث عن طرق تمكننا من تصميم الفعاليات باستخدام تكنولوجيا البيانات الموصولة Linked Data technologies. كما يستطيع المبرمجون اختيار لغة البرمجة، سواء أكانت ActionScript أو بايثون Python أو بيرل Perl، لمطابقة أو تحليل أو انتقاء مجموعة البيانات التي نعمل عليها. وتُستخدم لغة بيرل أحياناً في النشر.

نستخدم خرائط جوجل وبينغ وجوجل إيرث إلى جانب خرائط ArcMAP من Ersi لاستكشاف البيانات الجغرافية وتوضيحها مرئياً.

وللرسوم البيانية نستخدم باقة برامج Adobe Suite التي تتضمن After Effects وIllustrator وPhotoshop وFlash، على الرغم من أننا نادراً ما ننشر ملفات فلاش على الموقع الإلكتروني هذه الأيام مثل جافاسكريبت JavaScript، وخاصة JQuery. وهناك مكتبات جافاسكريبت أخرى مثل Highcharts وRaphael وD3 تواصل تلبية متطلباتنا للتوضيح المرئي للبيانات.

— بيلا هاريل وأندرو ليمدورفر، بي. بي. سي

كيفية عمل فريق تطبيقات الأخبار في صحيفة شيكاغو تريبيون

يتكون فريق تطبيقات الأخبار في صحيفة شيكاغو تريبيون من مجموعة مبرمجين منتقين وموزعين داخل غرفة الأخبار. نعمل بشكل وثيق مع المحررين والصحافيين للمساعدة في: (1) البحث في الأخبار ونقلها، (2) شرح الأخبار بشكل مصور على الإنترنت، (3) بناء مصادر معلومات دائمة على الشبكة لأهل شيكاغو الطيبين.

من المهم أن نجلس في غرفة الأخبار. وعادة ما نجد العمل أفضل من خلال الحديث وجها لوجه مع الصحافيين. وهم يعلمون أننا نسعد بمساعدتهم في استخراج البيانات من موقع حكومي معقد أو تحليل ملفات PDF أو تحويل ما لا يحتوي على بيانات إلى شيء يمكن تحليله. وهذا بمثابة العنصر الجاذب بالنسبة لفريقنا، وهكذا يمكننا معرفة مشاريع البيانات المحتملة منذ البداية.

وعلى عكس العديد من الفرق العاملة في هذا المجال، وظف فريقنا على يد تقنيين كانت الصحافة بالنسبة لهم مهنة جديدة. حصل بعضنا على شهادة الماجستير في الصحافة بعد عدة سنوات من كتابة الكود (الترميز) لأغراض تجارية، وتمت استعارة آخرين من جماعة الحكومة المفتوحة.

يتميز عملنا بسرعة البديهة وخفة الحركة. وللتأكد من أننا دائماً نعمل بانسجام وعلى وفاق في المكان، نبدأ كل صباح باجتماع سريع مدته خمس دقائق. وعادة ما نتوزع في مجموعات ثنائية، فعمل كل اثنين من المطوّرَين على جهاز واحد عادة ما يكون مثمراً أكثر مما إذا كانا يعملان على جهازين منفصلين. ومعظم المشاريع لا تستغرق أكثر من أسبوع لتكتمل؛ ولكن في المشاريع الأكبر، نعمل في روتين أسبوعي ونعرض عملنا على المعنيين أسبوعياً (غالباً ما يكونون من الصحافيين والمحررين). 'افشل سريعا' هو شعارنا، فإذا كنت مخطئا فيما تفعله، ينبغي أن تعلم هذا في أسرع وقت ممكن، خاصة إذا كنت تكتب الكود (ترميز البيانات)، وقبل حلول الموعد النهائي!

هناك جانب إيجابي هائل للعمل الروتيني المتكرر المحدد بموعد نهائي: فنحن نجدد أدواتنا باستمرار. كل أسبوع ننهي العمل سريعاً على تطبيق أو اثنين؛ ثم، وعلى خلاف صناع البرامج العاديين، يمكننا أن نضعه جانباً ونمضي قدما للعمل على المشروع التالي. هذه متعة نتشارك فيها مع الصحافيين، فكل أسبوع نتعلم أموراً جديدة.



نموذج 2–4. فريق تطبيقات الأخبار في صحيفة شيكاغو تريبيون (تصوير: هيذر بيلينجز)

تأتي جميع أفكار التطبيقات من الصحافيين والمحررين في غرفة الأخبار. وهذا في اعتقادي ما يميزنا عن فرق التطبيقات في غرف الأخبار الأخرى، والذين عادة ما يطرحون أفكارهم الخاصة بهم. لقد بنينا علاقات شخصية ومهنية في غرفة الأخبار، ويعلم الجميع أنهم يأتون إلينا عندما تكون لديهم بيانات.

معظم عملنا في غرفة الأخبار يصب في دعم الصحافيين. فنحن نساعدهم في التنقيب عن البيانات والبحث في ثناياها وتحويل ملفات PDF إلى جداول بيانية واستخراج البيانات من المواقع الإلكترونية وما إلى ذلك. هذه خدمة نسعد

بتقديمها لأنها تطلعنا مبكراً على العمل المتعلق بالبيانات الذي يجري في غرفة الأخبار. جزء من هذا العمل يتحول إلى تطبيق للأخبار: خريطة أو جدول أو أحياناً موقع على نطاق أوسع.

كنا فيما مضى نضع رابط التطبيق في صفحة الخبر المكتوب، ولكن لم ينتج عن هذا حركة مشاهدة عالية. أما هذه الأيام فنضع التطبيقات قرب أعلى صفحة الموقع الإلكتروني، ويوجد رابط للخبر في التطبيق، مما يناسب كليهما. هناك قسم لعملنا في الموقع الإلكتروني (http://www.chicagotribune.com/news/data/)، ولكن ليس عليه حركة مشاهدة مرتفعة. وهذا ليس مفاجئاً، فعبارة 'أنا أريد بعض البيانات اليوم!' ليست العبارة الأكثر تكرارا كنموذج للاستخدام.

يسُرّنا دائما ارتفاع معدلات مشاهدة الصفحات، ونحب أيضاً تقدير زملائنا لنا، ولكن هذا ليس كافياً. يجب أن يكون الدافع هو التأثير على حياة الناس والقانون ومساءلة السياسيين وما إلى ذلك. سيتناول الخبر المكتوب الموضوع الرائج وسيضفي عليه الطابع الإنساني من خلال بعض الطرائف. ولكن ماذا سيفعل القارئ بعد الفراغ من قراءة الخبر؟ هل عائلته في أمان؟ هل يحصل أطفاله على تعليم جيد؟ يتجلى عملنا في مساعدة القارئ على إيجاد خبر يخصه من خلال البيانات. بعض الأمثلة على العمل الشخصي والمؤثر الذي قمنا به تتضمن تقاريرنا عن السلامة في دور الرعاية (http://nursinghomes.apps.chicagotribune.com) وتطبيقات الشهادات الدراسية (http://schools.chicagotribune.com/).

— بريان بوير، صحيفة شيكاغو تريبيون

خلف الكواليس في مدونة البيانات لدى صحيفة الغارديان

عندما أطلقنا مدونة البيانات، لم تكن لدينا أدنى فكرة عمن سيهتم بالبيانات الخام والإحصاءات والتوضيحات المرئية. وكما قال أحد كبار الموظفين في المكتب: 'لا أرى سببا يدعو أحداً ما للاهتمام بهذا؟'

كان من المفترض أن تكون مدونة بيانات الغارديان (-http://www.guard

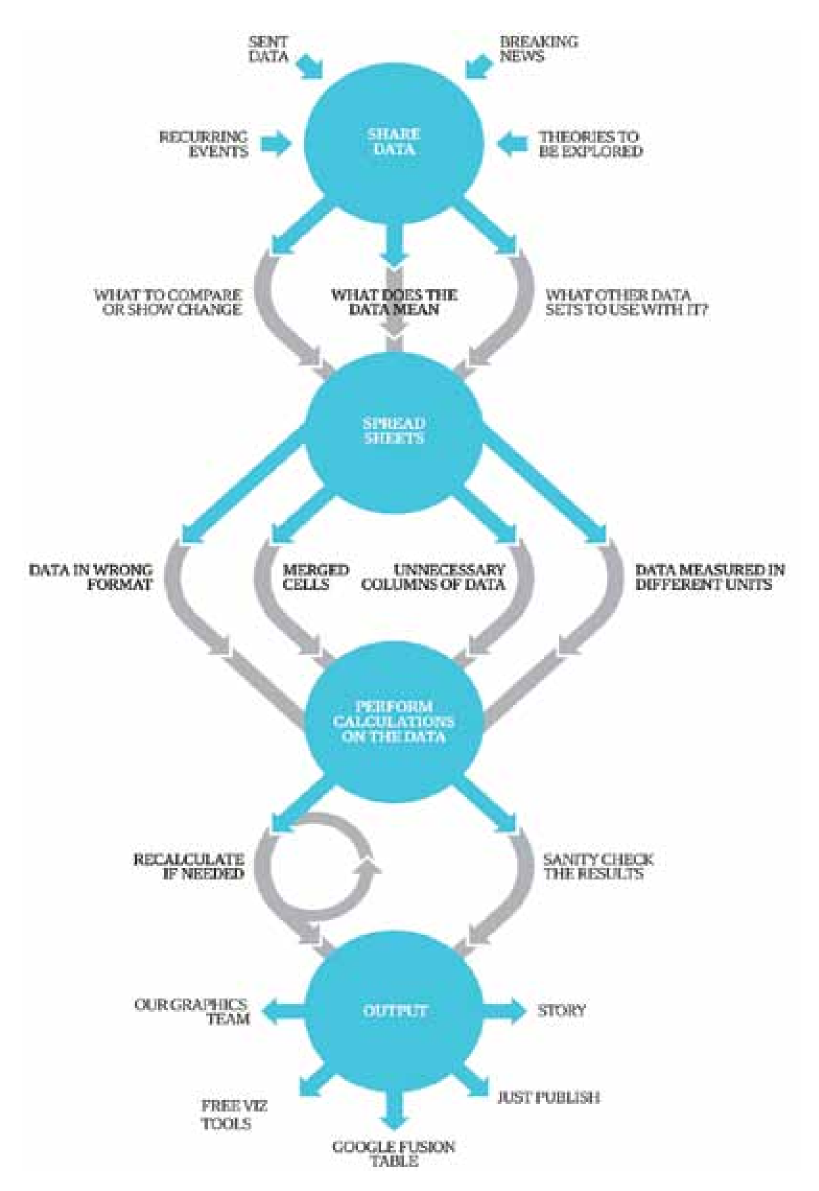
ian.co.uk/datablog) التي أقوم بتحريرها مدونة صغيرة تقدم مجموعات البيانات الكاملة وراء أخبارنا. لكنها تتكون الآن من صفحة أمامية (http://guardian.co.uk/data)، وبحث في بيانات حكومات العالم وبيانات التنمية الدولية، وإيضاحات مرئية للبيانات من مصممي الغرافيك في الغارديان ومن أنحاء شبكة الإنترنت، وأدوات لاستكشاف بيانات الإنفاق العام. فنحن نستخدم يومياً جداول جوجل البيانية لنشر البيانات الكاملة وراء عملنا، ونوضحها مرئياً ونحللها، ثم نستخدمها لتوفير الأخبار للصحيفة وللموقع الإلكتروني على السواء.

وكمحرر أخبار وصحفي يعمل مع الرسوم والأشكال البيانية، كان تجميع مجموعات بيانات جديدة ومعالجتها من أجل فهم أخبار اليوم يعتبر امتداداً منطقياً للعمل الذي أقوم به بالفعل. كانت هذه بضع سنوات استثنائية بالنسبة للبيانات العامة. فقد فتح أوباما خزانات بيانات الحكومة الأمريكية كأول قانون تشريعي له، ثم حذت حذوه مواقع البيانات الحكومية حول العالم: أستراليا ونيوزيلندا وموقع الحكومة البريطانية Data.gov.uk

كانت هناك فضيحة نفقات نواب البرلمان، وهي أكثر أخبار صحافة البيانات مفاجأة في بريطانيا، وقد أدت تداعياتها إلى التزام وستمنستر الآن بنشر كم هائل من البيانات سنوياً.

وأجرينا انتخابات عامة التزم فيها كل من الأحزاب السياسية الرئيسية بشفافية البيانات وفتح خزانات بياناتنا الخاصة أمام العالم. كما خصصت الصحف مساحات قيّمة لنشر قاعدة البيانات COINS الخاصة بالخزانة العامة.

وفي الوقت ذاته، بينما تستمر شبكة الإنترنت في ضخ المزيد من البيانات، يزاد اهتمام القراء حول العالم بالحقائق الخام وراء الأخبار أكثر من أي وقت مضى. وعندما أطلقنا مدونة البيانات، كنا نظن أن جمهورنا سيكون من المطوّرين الذين يبنون التطبيقات. ولكن في الواقع، كان جمهورنا يشمل هؤلاء الذين يريدون معرفة المزيد عن انبعاثات غاز الكربون أو معدلات الهجرة من أوروبا الشرقية أو توزيع الوفيات في أفغانستان، أو حتى عدد المرات التي ذكرت فيها كلمة 'حب' في أغاني فرقة البيتلز (613).



نموذج 2–5. التوضيح المرئي لعملية إنتاج مدونة البيانات لدى صحيفة الغارديان (الغارديان)

وبصورة تصاعدية عكست مدونة البيانات تدريجياً الأخبار التي نتناولها، بل وأضافت لها. وقد لجأنا إلى أسلوب التكليف الجماعي للتعامل مع 458,000 وثيقة تتعلق بنفقات نواب البرلمان وحللنا البيانات المفصلة التي قدمها النواب. كما ساعدنا المستخدمين على استكشاف قواعد بيانات إنفاق الخزانة العامة، ونشرنا

البيانات التي كانت وراء الخبر.

ولكن التحول الحقيقي في صحافة البيانات حدث في ربيع عام 2010، وقد بدأ بجدول بياني واحد: 92,201 صفا من البيانات، يحتوي كل منها على تحليل مفصل لحدث عسكري في أفغانستان. كانت هذه سجلات الحرب الخاصة بموقع ويكيليكس، أو على الأقل الجزء الأول منها. كان هناك جزءان لاحقان: سجلات العراق والبرقيات. كان المصطلح الرسمي لأول جزأين هو SIGACTS: أي قاعدة بيانات أهم الإجراءات العسكرية الأمريكية.

تعتمد مؤسسات الأخبار أساساً على الموقع الجغرافي والقرب من مكتب الأخبار. فإذا كنت قريبا، سيسهل عليك أن تقترح الأخبار وأن تكون جزءاً من العملية، أما إن كنت عكس ذلك، فالبعيد عن العين يكون فعلاً بعيداً عن العقل. فقبل ظهور موقع ويكيليكس كنا في طابق آخر، مع قسم الغرافيكس. أما بعد ظهور الموقع، فقد أصبحنا في نفس الطابق إلى جانب مكتب الأخبار. ولذا أصبح من الأسهل علينا اقتراح الأفكار على المكتب، وأصبح من الأسهل للصحافيين من الجانب المقابل في غرفة الأخبار أن يلجأوا إلينا طلبا لمساعدتهم في الأخبار.

منذ زمن ليس ببعيد كان الصحافيون هم حراس البيانات الرسمية. كنا نكتب الأخبار بشأن الأرقام ثم ننشرها للجمهور الذي لم يكن مهتماً بالإحصاءات الخام. كانت فكرة السماح بنشر معلوماتنا الخام في صحفنا أمرا كريها.

أما الآن فقد تغير هذا كلياً. فتحول دورنا إلى مفسرين لنساعد الناس على فهم البيانات، حتى إننا أصبحنا ننشرها لمجرد أنها مثيرة للاهتمام.

ولكن الأرقام بدون تحليل تبقى مجرد أرقام، وهنا يأتي دورنا. حين زعم رئيس الوزراء البريطاني أن أحداث الشغب في أغسطس 2011 لم تكن بسبب الفقر، استطعنا وضع خريطة لعناوين مثيري الشغب مع مؤشرات للفقر حتى نوضّح حقيقة مزاعمه.

هناك عملية تتغير باستمرار وهي التي تحرك أخبار صحافة البيانات، حيث نستخدم أدوات وأساليب جديدة. ويرى البعض أن الحل هو أن نصبح نوعاً من المبرمجين الخارقين، نكتب المدونات وننخرط في لغة SQL. يمكنك أن تختار هذا النهج، ولكننا في الكثير من عملنا نستخدم برنامج إكسل فقط.

أولاً، نحدد مكان البيانات أو نتلقاها من مختلف المصادر مثل الأخبار العاجلة والبيانات الحكومية وأبحاث الصحافيين وما إلى ذلك. ثم نبدأ البحث فيما يمكننا فعله بتلك البيانات؛ هل نحتاج لدمجها مع مجموعة بيانات أخرى؟ كيف يمكننا توضيح التغيرات عبر الزمن؟ عادة ما يجب أن تكون تلك الجداول البيانية مرتبة على نحو جيد للغاية، فكل هذه الأعمدة والخانات الغريبة حقاً لا تساعد على سهولة استيعابها. وهذا على اعتبار أنها ليست ملفات من نوع PDF، وهي أسوأ صيغة للبيانات عرفتها البشرية.

عادة ما تأتي البيانات الرسمية مع الرمز الرسمي مضافاً إليها، ولكل مدرسة أو مستشفى أو دائرة انتخابية أو سلطة محلية رمز تعريفي مميز.

وللبلدان أيضاً رموزها (فرمز المملكة المتحدة مثلاً هو GB). وهي مفيدة لأنك قد تريد دمج مجموعات البيانات أو مقارنتها ببعضها، وقد يدهشك كم الاختلافات الإملائية وترتيب الكلمات التي يمكن أن تعترض طريقك. فهناك مثلاً بورما وميانمار.. اسمان لبلد واحد، أو مقاطعة باسم "فايت" في الولايات المتحدة (يوجد 11 فايت في الولايات من جورجيا إلى ويست فيرجينيا). والرموز تسمح لنا بالتفريق بين الأسماء المتشابهة.

وفي نهاية تلك العملية يأتي الناتج: هل سيكون خبراً أم رسماً بيانياً أم توضيحا مرئياً؟ وما هي الأدوات التي سنستخدمها؟ أفضل أدواتنا هي الأدوات البسيطة المجانية التي تمكننا من عمل شيء سريع. أما الرسوم البيانية الأكثر تعقيدا فينتجها فريق المطوّرين لدينا.

هذا يعني أننا كثيراً ما نستخدم جداول جوجل للرسوم البيانية الصغيرة، أو جداول جوجل فيوجن لرسم الخرائط بسرعة وبسهولة.

قد يبدو هذا جديداً، ولكنه ليس كذلك.

في العدد الأول من صحيفة مانشستر غارديان (السبت 5 مايو 1821)، كانت الأخبار في الصفحة الخلفية، مثل أي صحيفة في ذلك الوقت. وكان الجزء الأول من الصفحة الأمامية عبارة عن إعلان لكلب من فصيلة لابرادور مفقود.

ووسط الأخبار والمقتطفات الشعرية، كان ثلث الصفحة الخلفية مخصصا للحقائق. كان جدولاً مفصلاً عن تكاليف المدارس في المنطقة 'لم ينشر للجمهور

من قبل' كما كتب 'ن. هـ'.

أراد "ن. هـ." أن ينشر بياناته لأنه إن لم يفعل ذلك ستُترك البيانات لموظفين غير مدربين لينشروها. كان دافعه هو أن "مثل تلك المعلومات قيّمة بما تحتويه، لأنه بدون معرفة مدى سيادة التعليم، ستضلل الآراء حول حالة تقدم المجتمع ومستقبله." وبعبارة أخرى، إن لم يعلم الناس بما يجري، كيف يمكن للمجتمع أن يتحسن؟

لا يمكنني الآن أن أذكر مبرراً أفضل لما نحاول القيام به، فما كان خبراً في الصفحة الأخيرة يمكنه الآن أن يكون من ضمن أخبار الصفحة الرئيسية.

—— سايمون روجرز، صحيفة الغارديان

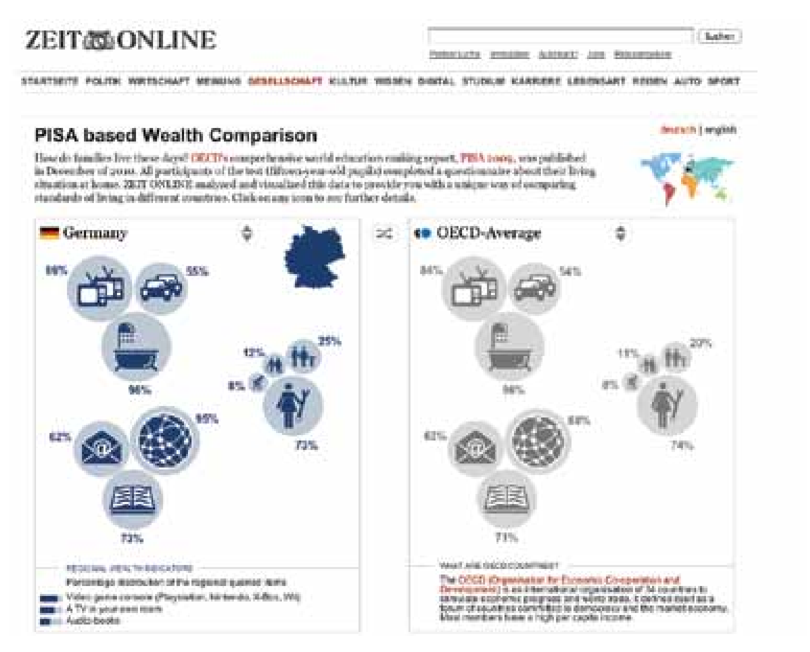
صحافة البيانات في صحيفة دي تسايت الإلكترونية

يعد مشروع مقارنة الثروات القائم على أساس برنامج تقييم الطالب الدولي (http://bit.ly/Pisa_Wealth) توضيحا مرئيا تفاعليا يتيح عقد مقارنة بين مستويات المعيشة في بلاد مختلفة. يستخدم هذا المشروع بيانات من التقرير المفصل لترتيب التعليم العالمي الصادر عن منظمة التعاون الاقتصادي والتنمية OECD تحت عنوان PISA 2009 (http://bit.ly/Pisa_2009) الذي نُشر في ديسمبر 2010. يعتمد التقرير على استطلاع رأي يسأل طلاباً تبلغ أعمارهم 15 عاما عن مستوى معيشتهم في المنزل.

كان الغرض من الاستطلاع تحليل تلك البيانات وتوضيحها مرئياً لتقديم طريقة فريدة لمقارنة مستويات المعيشة في البلاد المختلفة.

أولاً، حدد فريق المحررين لدينا الحقائق التي تبدو مفيدة لمقارنة مستويات المعيشة التي يجب توضيحها مرئياً، وتضمنت:

- الثروة (عدد ما يملكونه من أجهزة التلفزيون والسيارات وعدد الحمامات في كل منزل)
- الحالة العائلية (ما إذا كان هناك جد أو جدة يعيشون مع العائلة، نسبة العائلات ذات الطفل الواحد، بطالة الأبوين والحالة الوظيفية للأم)
- القدرة على الوصول إلى مصادر المعرفة (الإنترنت في المنزل، عدد مرات



نموذج 2–6. مشروع مقارنة الثروات المبني على برنامج تقييم الطالب الدولي (صحيفة دي تسايت الإلكترونية/ النسخة الإنجليزية)

استخدام البريد الإلكتروني وعدد الكتب التي يملكونها)

- ثلاثة مؤشرات إضافية لمستوى التنمية في كل بلد.

وبمساعدة فريق التصميم الداخلي، تمت ترجمة تلك الحقائق إلى أيقونات مبسطة لا تحتاج إلى شرح. كما تم وضع تصميم لجعل المقارنة بين البلاد المختلفة ممكنة عن طريق النظر إليها كأنها أوراق لعب.

بعد ذلك اتصلنا بالشبكة الألمانية للبيانات المفتوحة (http://opendata-network.org/) لنبحث عن مطوّرين يمكنهم مساعدتنا في المشروع. وقد قدمت لنا هذه الجماعة النشطة السيد غريغور آيش - وهو مصمم معلومات موهوب للغاية - ليكتب كود التطبيق الذي من شأنه أن يحقق أحلامنا (بدون استخدام فلاش، الأمر الذي كان في غاية الأهمية بالنسبة لنا!). صمم غريغور توضيحاً مرئياً تفاعلياً ذا جودة عالية جداً على شكل فقاعة جميلة بناء على مكتبة Raphaël-Javascript (http://raphaeljs.com/).

نتج عن تعاوننا هذا تصميم تفاعلي ناجح للغاية، وحصل على نسبة مشاهدة

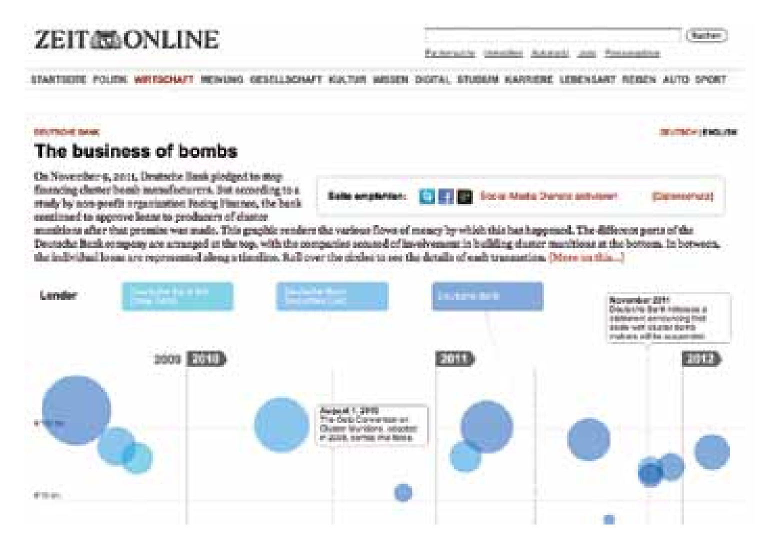
عالية. فقد أصبح من السهل مقارنة أي بلدين مما جعله مفيداً كأداة مرجعية. وهذا يعني أنه بإمكاننا إعادة تطبيقه في عملنا التحريري اليومي. فعلى سبيل المثال، إذا كنا نغطي خبراً يتعلق بمستوى المعيشة في إندونيسيا، يمكننا بسرعة وبسهولة أن نضع رسماً بيانياً يقارن بين مستويات المعيشة في إندونيسيا وألمانيا (http://bit.ly/Pisa_Indonesia_Germany). وقد كان نقل تلك المعرفة إلى فريقنا استثماراً رائعاً للمشاريع المستقبلية.

في صحيفة دي تسايت الإلكترونية، وجدنا أن مشاريع صحافة البيانات الخاصة بنا (http://www.zeit.de/datenjournalismus) قد جلبت نسبة مشاهدة عالية، وساعدتنا على إشراك الجمهور بطرق جديدة. فمثلاً، كانت هناك تغطية واسعة للموقف في ألمانيا بتأثير حادث المفاعل النووي في فوكوشيما بعد التسونامي في اليابان. فبعد تسرب المواد المشعة من المفاعل، تم إخلاء جميع السكان حتى مسافة 30 كيلومترا منه. كان الناس يقرأون ويرون الكثير بشأن عمليات الإخلاء. وقد وجدت صحيفة دي تسايت الإلكترونية طريقة لشرح تأثير ذلك لجمهورنا الألماني. سألنا: كم من الناس يسكنون قرب مفاعل نووي في ألمانيا؟ كم من الناس يسكنون في محيط دائرة نصف قطرها 30 كيلومترا حول المفاعل؟ تظهر الخريطة أعداد الناس الذين يجب أن يتم إخلاؤهم في موقف مماثل في ألمانيا (http://bit.ly/near–nuclear). النتيجة: نسبة مشاهدة عالية جداً، بل وانتشر المشروع على شبكات التواصل الاجتماعي. يمكن تعديل مشاريع صحافة البيانات بسهولة نسبية إلى لغات أخرى. وقد وضعنا نسخة إنجليزية في المناطق القريبة من المفاعلات النووية في الولايات المتحدة، مما حقق نسبة مشاهدة عالية. تسعى مؤسسات الأخبار إلى أن تُعرف وسط قرائها بأنها مصادر رسمية وموثوقة. وقد وجدنا أن مشاريع صحافة البيانات إلى جانب تمكين القراء من البحث وإعادة استخدام البيانات الخام قد حققت لنا درجة عالية من المصداقية.

على مدار عامين في قسم البحث والتطوير، دعا وولفجانج بلو، رئيس تحرير صحيفة دي تسايت الإلكترونية، إلى استخدام أوسع صحافة البيانات في سرد الأخبار. يذكر أن الشفافية والمصداقية وإشراك المستخدم كلها أجزاء هامة من فلسفتنا. ولذا تمثل صحافة البيانات جزءاً طبيعياً من عملنا الحالي والمستقبلي.

يمكن للتوضيحات المرئية أن تعطي قيمة لتلقي الخبر، وهي طريقة جذابة لفريق المحررين بأكمله لعرض ما لديهم من محتوى.

على سبيل المثال، في 9 نوفمبر 2011 تعهد البنك الألماني Deutsche Bank بوقف تمويل صناع القنابل العنقودية. ولكن وفقاً لدراسة أجرتها منظمة Facing Finance غير الربحية، واصل البنك منح قروض لمنتجي الذخائر العنقودية. وقد نجحنا في أن نُظهر لقرائنا من خلال توضيحنا المرئي (/http://zeit.de/wirtschaft cluster–munition) المبني على البيانات تدفقات الأموال المختلفة. في الأعلى تجد الأقسام المختلفة من مؤسسة البنك الألماني، وفي الأسفل تجد الشركات المتهمة بالضلوع في صنع الذخائر العنقودية. وفيما بينهما، تجد القروض الفردية مرتبة ترتيباً زمنياً. وعند التنقل بين الدوائر، ترى تفاصيل كل عملية. بالطبع كان يمكن سرد الخبر في مقال مكتوب. ولكن التوضيحات المرئية تمكن قراءنا من فهم واستكشاف التبعية المالية بطريقة أكثر سلاسة.



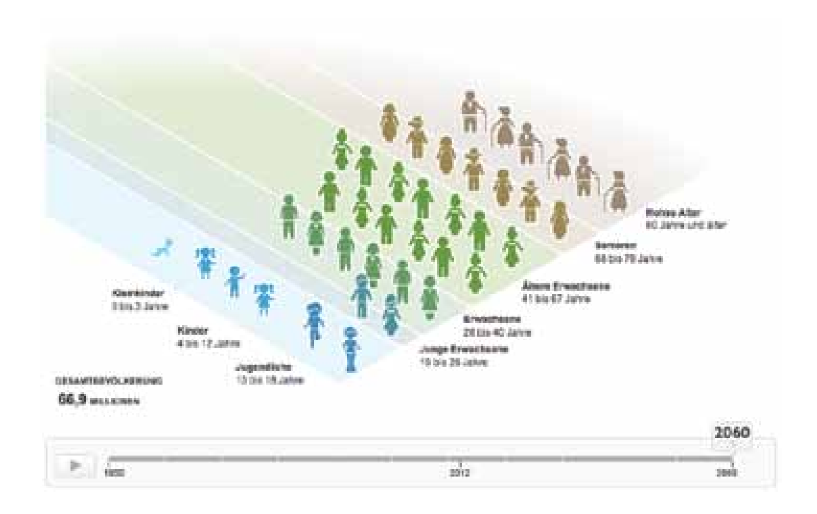
نموذج 2–7. تجارة القنابل (صحيفة دي تسايت الإلكترونية/ النسخة الإنجليزية)

وإلى مثال آخر: نشر مكتب الإحصاءات الفيدرالي في ألمانيا (//:http www.destatis.de /EN /Homepage.html) مجموعة بيانات رائعة عن إحصاءات هامة بالنسبة لألمانيا، تضم تصميم سيناريوهات ديمغرافية مختلفة حتى

عام 2060 (http://bit.ly/German_Federal_Statistics)، بخلاف الطريقة المعتادة لتقديم ذلك برسم هرم سكاني، مثل ذلك الذي كان في موقع وكالة الإحصاءات الفيدرالية (http://www.destatis.de/bevoelkerungspyra-mide/).

وبالتعاون مع زملائنا من القسم العلمي، حاولنا توفير طريقة أفضل لقرائنا ليستكشفوا البيانات الديمغرافية المتوقعة في المجتمع الألماني في المستقبل. وفي توضيحنا المرئي (http://www.zeit.de/wissen/altersstruktur)، نقدم مجموعة ممثلة بالإحصاءات ومكونة من 40 شخصا من أعمار مختلفة خلال الأعوام ما بين 1950 إلى 2060. وقد نظمناهم في ثماني مجموعات مختلفة، تبدو كصورة جماعية للمجتمع الألماني في مراحل مختلفة من الزمن. إذا أوضحنا نفس البيانات مرئياً في الهرم السكاني المعتاد، قد نحصل على نظرة مجردة للوضع، أما إذا كانت مجموعة من الأطفال والشباب والبالغين وكبار السن، فإن هذا يعني أن قراءنا يمكنهم التعرف على البيانات بشكل أسهل واكثر تفصيلا. يمكنك بكل بساطة أن تضغط زر التشغيل لتبدأ رحلة عبر أحد عشر عقدا من الزمان. يمكنك أيضاً أن تدخل سنة ميلادك وجنسك لتصبح جزءاً من الصورة الجماعية: لترى رحلتك الديمغرافية عبر الزمن ومتوسط عمرك المتبقي.

— ساشا فينور، صحيفة دي تسايت الإلكترونية



نموذج 2–8. التوضيح المرئي للبيانات الديمغرافية
(صحيفة دي تسايت الإلكترونية/ النسخة الألمانية)

كيف توظف مبرمج كمبيوتر

كثيراً ما يسألني الصحافيون 'كيف أجلب مبرمجاً ليساعدني في مشروعي؟' لا تنخدع في الاعتقاد بأنها عملية من طرف واحد، فعادة ما يتوق المبرمجون المدنيون ومعالجو البيانات أيضاً إلى الاتصال بالصحافيين.

لا شك في أن الصحافيين مستخدمون ماهرون للأدوات والخدمات المبنية على البيانات. ومن وجهة نظر المطورين، يعتبر تفكير الصحافيين إبداًعياً لأنهم يستخدمون أدوات البيانات في سياق لم يخطر على بال المطوّرين من قبل (فإعطاء الملاحظات أمر لا يقدر بثمن!). كما يساعدون في وضع السياق ويثيرون الاهتمام حول المشاريع لجعلها هامة. إذن فعلاقتهم علاقة تكافلية.

ومن حسن الحظ، يعني هذا أنك سواء كنت تسعى إلى توظيف مبرمج أو تبحث عن أوجه للتعاون ضمن ميزانية محدودة، سيكون من الأرجح أن تجد شخصا ما يحرص على مساعدتك باهتمام.

إذن فكيف تجدهم؟ يقول آرون بيلهوفر من صحيفة نيويورك تايمز:

"قد تجد أن المؤسسة التي تعمل بها لديها بالفعل أشخاص يملكون المهارات المطلوبة، ولكنهم ليسوا بالضرورة موجودين معك في غرفة الأخبار. تجوّل قليلاً، قم بزيارة أقسام التكنولوجيا والمعلومات ومن الأرجح أن تجد غايتك. ومن المهم أيضاً أن تقدّر ثقافة كاتبي المدونات، حاول أن تجد شخصا لديه مثل هذا الكمبيوتر في نموذج 2–9 وتكون حينئذٍ قد وجدت ضالتك."

فيما يلي المزيد من الأفكار:

انشر إعلانات على مواقع التوظيف

تعرف على المواقع الموجهة إلى المطوّرين الذين يعملون بلغات برمجة مختلفة وانشر إعلاناتك فيها. على سبيل المثال، لوحة وظائف بايثون (http://www.python.org/community/jobs).



نموذج 2–9. وسام الشرف: يسهل عادة العثور على المبرمجين (تصوير: لوسي تشامبرز)

تواصل مع قوائم البريد الإلكتروني ذات الصلة

مثل قوائم NICAR-L (http://bit.ly/nicar-subscribe) وصحافة البيانات (http://bit.ly/ddj-list).

تواصل مع المؤسسات ذات الصلة

فمثلاً، إن كنت تريد أن تصفي البيانات أو تستخرجها من موقع إلكتروني ما، اتصل بمؤسسة مثل Scraperwiki (http://scraperwiki.com/) التي لديها قائمة رائعة لعناوين كاتبي الكود الموثوقين والمستعدين للعمل.

انضم إلى المجموعات أو الشبكات ذات الصلة

ابحث عن المبادرات مثل ملتقى الصحافيين والمبرمجين Hacks/Hackers (http://hackshackers.com/) الذي يجمع الصحافيين ومحبي التكنولوجيا معاً. وتظهر مجموعات Hacks/Hackers الآن في جميع أنحاء العالم. يمكنك أيضاً أن تحاول نشر إعلاناتك في نشرتهم الوظيفية (http://bit.ly/hacks-hackers-jobs).

الجماعات المحلية المختصة

يمكنك إجراء بحث سريع عن تخصص معين في منطقتك (مثل: 'جافاسكريبت' + 'لندن'). وهناك مواقع مثل Meetup.com قد تكون نقطة بداية جيدة جداً.

ملتقيات المبرمجين والمسابقات

سواء أكانت هناك مكافأة مالية أم لم تكن، فإن مسابقات التطبيقات والتوضيحات المرئية وأيام التطوير عادة ما تكون أرضاً خصبة للتعاون وبناء العلاقات.

اسأل خبراء الكمبيوتر!

عادة ما يقضي خبراء الكمبيوتر وقتهم مع أقرانهم. وستجد دائماً أن الكلام الشفهي وسيلة جيدة للعثور على أشخاص للعمل معهم.

— لوسي تشامبرز، مؤسسة المعرفة المفتوحة

مهارات المبرمجين

عندما تعثر على مبرمج، كيف تتأكد من أنه يجيد عمله؟ سألنا 'ألاستر دانت' من صحيفة الغارديان:

يمكنه كتابة جميع أنواع الشفرة (الكود)

عند التعامل مع المواعيد النهائية، يستحسن أن يكون المبرمج متعدد المهارات بدلاً من أن تكون لديه مهارة واحدة. فتطبيقات الأخبار تتطلب معالجة البيانات وتصميم الرسوم البيانية الديناميكية والعمل تحت الضغط.

يمكنه رؤية الصورة الكاملة

يفضّل التفكير الشمولي من منظور واسع للسرد على التفاصيل الفنية. حاول معرفة ما إذا كان الشخص المعني يسره العمل بجوار المصمم.

يمكنه السرد بشكل جيد

يتطلب السرد ترتيب الأشياء زمنياً ومكانياً. اعرف المشروع الذي يفخر به واطلب منه أن يحكي لك عن كيفية تنفيذه، سيكشف هذا الكثير عن قدراته على

التواصل وفهمه التقني للأمور.

يمكنه مناقشة الأمور

يستلزم بناء الأشياء سريعا عمل مزيج من الفرق سعيا لتحقيق أهداف مشتركة. يجب على كل مشارك أن يحترم زملاءه وأن يكون مستعدا للتفاوض. وعادة ما تتطلب العقبات المفاجئة إعادة تخطيط وتنازلات جماعية سريعة.

يمكنه أن يعلم نفسه بنفسه

تتحرك التكنولوجيا بسرعة. وتستلزم ملاحقتها الكثير من الجهد. وبعد أن التقيت بالعديد من المطوّرين الجيدين من مختلف الخلفيات، وجدت أن الصفة المشتركة بينهم هي الاستعداد لتعلم الجديد والضروري بسرعة.

—— لوسي تشامبرز، مؤسسة المعرفة المفتوحة، في مقابلة مع ألاستر دانت، مدير التقنيين التفاعليين في صحيفة الغارديان.

كيف تعثر على المطوّر الذي تتمناه

الفرق بين المطوّر الجيد والمطوّر الممتاز ليس بسيطاً، بل هائل. فاختيار الموظفين الجيدين أمر هام للغاية، ولكنه أيضاً – للأسف – في غاية الصعوبة. فالتدقيق في المرشحين صعب بما يكفي إن لم تكن مديراً تقنيا ذا خبرة. أضف إلى ذلك الرواتب التي تستطيع مؤسسات الأخبار توفيرها، وسيكون أمامك تحد كبير.

في صحيفة التريبيون، نوظّف وفقاً لمعيارين: الجانب المعنوي والجانب التقني. الجانب المعنوي هو الآتي: إن الصحافة عنصر أساسي لإقامة ديمقراطية فعالة. إذا عملت هنا يمكنك تغيير العالم. أما على الجانب التقني، فنحن نشجع على ما يمكنك تعلمه. فمشاريعنا صغيرة وسريعة ومكررة، ويحتوي كل مشروع على مجموعة جديدة من الأدوات ولغة جديدة وموضوع جديد (كنظام إطفاء الحرائق أو نظام المعاشات التقاعدية) التي لا بد أن تتعلمها. وغرفة الأخبار بمثابة بوتقة اختبار. لم يسبق لي أن ترأست فريقاً تعلم بالسرعة التي تعلم بهما فريقنا.

أما بالنسبة لمكان البحث، فقد حالفنا الحظ كثيراً في العثور على مبرمجين

ذوى كفاءة عالية ضمن مجتمع الحكومة المفتوحة. إن قائمة The Sunlight Labs البريدية هي المكان الذي يتسكع فيه خبراء الكمبيوتر المتميزون الذين يعملون نهاراً في وظائف سيئة. هناك مصدر محتمل آخر وهو Code for America حيث تظهر منه سنوياً مجموعة من الزملاء الباحثين عن مشروعهم الكبير القادم. وكميزة إضافية، تجد أنهم شديدو الصرامة في إجراء المقابلات. إذن فقد قاموا بالفعل بعملية التدقيق من أجلك. واليوم، يظهر الصحافيون المهتمون بالبرمجة من كليات الصحافة. ورغم خبراتهم القليلة، إلا أنهم يتمتعون بإمكانات هائلة.

وأخيراً، لا يكفي أن توظف مطوّرين فقط، فأنت تحتاج إلى الإدارة التقنية. فالمطوّر وحده (وخاصة إذا تخرج حديثا من كلية الصحافة وليست لديه أي خبرة في المجال) سيتخذ العديد من القرارات السيئة. وحتى أفضل المبرمجين، إذا تُرك له القرار، سيختار العمل التقني الذي يثير اهتمامه هو بدلاً مما يهم جمهورك.

يمكنك أن تسمي هذا ما شئت: توظيف محرر لتطبيقات الأخبار أو مدير للمشاريع. ومثل الكُتاب، يحتاج المبرمجون إلى المحررين والموجهين وإلى شخص يشجعهم من أجل الالتزام بالموعد النهائي لإنجاز برنامج ما.

— بريان بوير، صحيفة شيكاغو تريبيون

الاستفادة من خبرات الآخرين من خلال ملتقى المبرمجين (الهاكثون)

في مارس 2010، قامت منظمة الثقافة الرقمية SETUP التي يقع مقرها في مدينة 'أوترخت' بالدعوة إلى فعالية تحت عنوان 'برمجة الصحافة' (http://setup.nl/content/hacking-journalism)، وتم تنظيمها للتشجيع على المزيد من التعاون بين المطوّرين والصحافيين.

يقول المبرمجون "نحن ننظم ملتقيات البرمجة لابتكار تطبيقات جيدة، ولكننا لا نستطيع تحديد الأخبار المثيرة للاهتمام ضمن البيانات. ما نبنيه لا أهمية اجتماعية له." بينما يقول الصحافيون "يمكننا التعرف على أهمية صحافة البيانات، ولكننا لا نملك المهارات التقنية اللازمة لنقوم بما نريده."



نموذج 2–10. الصحافيون والمطوّرون في ملتقى RegioHack (تصوير: هاينز هافينجا)

ولأننا نعمل في صحيفة محلية، لم يكن هناك ما يكفي من المال أو الحافز لتوظيف مبرمج في غرفة الأخبار. كانت صحافة البيانات لا تزال أمراً مجهولا بالنسبة للصحف الهولندية في ذلك الوقت.

كان نموذج ملتقى المبرمجين مثالياً. فقد عُقد في أجواء مريحة تبعث على التعاون، مع الكثير من البيتزا ومشروبات الطاقة. نظم ملتقى المبرمجين RegioHack (http://www.regiohack.nl/) وصحيفة دي ستينتور المحلية De Stentor (http://www.destentor.nl/) التي أعمل بها والصحيفة الشقيقة TC Tubantia (http://www.tctubantia.nl) وSaxion Hogescholen Enschede (http://saxion.nl) الذين وفروا مكان انعقاد اللقاء. وكان البرنامج كالتالي: يمكن للجميع التسجيل في ملتقى للمبرمجين مدته 30 ساعة. وقمنا نحن بتقديم الطعام والشراب. كنا نخطط لتسجيل 30 مشاركاً نوزعهم على ست مجموعات. وتركز تلك المجموعات على مواضيع مختلفة مثل الجريمة أو الصحة أو النقل أو السلامة أو التقدم في السن أو السلطة. وبالنسبة لنا، كانت الأهداف الثلاثة الرئيسية لهذا الحدث كالآتي:

العثور على الأخبار

لأن صحافة البيانات بالنسبة لنا أمر جديد ومجهول. والطريقة الوحيدة التي يمكننا بها إثبات فائدتها هي عبر الأخبار المحبوكة بمهارة. وقد خططنا لإنتاج ثلاث قصص إخبارية على الأقل.

جمع الناس

نحن، الصحافيين، لا نعلم كيف تعمل صحافة البيانات ولا ندعي العلم. وبجمع الصحافيين والطلاب والمبرمجين في غرفة واحدة لمدة 30 ساعة، نريد منهم أن يشاركوا بمعرفتهم وأفكارهم.

استضافة فعالية اجتماعية

لا تنظم الصحف العديد من الفعاليات الاجتماعية، ناهيك عن ملتقى المبرمجين. أردنا أن نختبر كيف يمكن لمثل هذا الحدث أن يحقق نتائج. وفي الواقع، كان يمكن للملتقى أن يؤدي إلى التوتر بسبب قضاء 30 ساعة مع الغرباء واستخدام الكثير من المصطلحات المتخصصة والتفكير ملياً في إجابات الأسئلة الأساسية والعمل خارج إطار المألوف الذي صرت مرتاحا لممارسته. ومن خلال إضفاء الصبغة الاجتماعية على اللقاء (هل تتذكر البيتزا ومشروبات الطاقة؟)، أردنا أن نخلق بيئة مريحة للصحافيين والمبرمجين ليتعاونوا بشكل فعّال.

قبل الفعالية، نشرت صحيفة TC Tubantia مقابلة مع أرملة شرطي كانت قد كتبت كتاباً عن سنوات عمل زوجها، وكان لديها أيضاً وثيقة احتفظ بها زوجها منذ عام 1945 تشمل جميع جرائم القتل المسجلة في الجزء الشرقي من هولندا. كنا سننشر تلك الوثيقة على موقعنا الإلكتروني، ولكن هذه المرة، وضعنا لوحة معلومات باستخدام برنامج Tableau (http://bit.ly/tableau-dashboard). كما نشرنا مدونة (http://bit.ly/regiohack-blog) حول كيفية قيامنا بذلك على موقع RegioHack الخاص بنا.

أثناء الملتقى، اقترحت إحدى المجموعات موضوع تطوير المدارس وتعمير منطقتنا. ومن خلال التوضيح المرئي للتوقعات المستقبلية (http://bit.ly/tableau-workbook)، تمكنّا من فهم أي المدن ستتعرض لمشاكل انخفاض معدلات الالتحاق بالمدارس بعد بضع سنوات. وانطلاقاً من هذه الرؤية، نشرنا مقالاً حول كيفية تأثير ذلك على المدارس في منطقتنا.

وقد بدأنا أيضاً العمل على مشروع طموح جداً يسمّى De Tweehonderd van Twente (200 من منطقة توينتي أو بالإنجليزية The Two Hundred of Twenty) لنحدد من كان الأكثر نفوذاً في منطقتنا ولنبني قاعدة بيانات بأكثر الشخصيات تأثيرا. وبالقيام بعملية حسابية تقربنا إلى بحث جوجل - من لديه أقوى العلاقات بالمؤسسات ذات النفوذ - سنكوّن قائمة بأكثر الشخصيات تأثيراً. قد يؤدي ذلك إلى سلسلة مقالات، ولكنه أيضاً يشكل أداة قوية للصحافيين. من لديه علاقات مع من؟ يمكنك أن تطرح أسئلة على قاعدة البيانات تلك وتستخدمها بشكل يومي. كما أن لقاعدة البيانات هذه قيمة ثقافية، فقد سأل الفنانون بالفعل ما إذا كان بإمكانهم استخدامها عند الانتهاء منها لصنع تركيبات فنية تفاعلية.

بعد ملتقى RegioHack، لاحظنا أن الصحافيين يعتبرون صحافة البيانات إضافة عملية إلى الصحافة التقليدية. وقد واصل زملائي استخدام الأساليب التي تعلمناها ذلك اليوم والإضافة إليها لابتكار مشاريع تقنية أكثر طموحاً مثل قاعدة بيانات التكاليف الإدارية للإسكان. وباستخدام تلك البيانات، رسمت خريطة تفاعلية باستخدام جداول فيوجن (http://bit.ly/stentor-map)، وطلبنا من قرائنا مثلاً أن يختبروا البيانات، وقمنا بفرز وتصنيف النتائج في نماذج جماعية هنا (http://bit.ly/scratchbook-crowdsourcing). وبعد العديد من الأسئلة حول كيفية رسمنا لخريطة باستخدام جداول فيوجن، قمت أيضاً بتسجيل مقطع فيديو تعليمي عن الموضوع (http://bit.ly/vermanen-video).

ماذا تعلمنا إذن؟ تعلمنا الكثير، ولكننا أيضاً واجهنا العديد من العقبات، ومن بينها حددنا العقبات الأربع التالية:



نموذج 2–11. مجتمعات جديدة تهتم بصحافة البيانات (تصوير: هاينز هافينجا)

من أين نبدأ: السؤال، أم البيانات؟

تتعطل جميع المشاريع تقريباً عند البحث عن معلومات. وفي أغلب الأحيان، نبدأ بطرح أسئلة صحفية. ولكن ماذا بعد؟ ما هي البيانات المتاحة؟ أين يمكننا العثور عليها؟ وحين نجدها، هل يمكننا استخدامها للإجابة على السؤال؟ في الظروف العادية يعلم الصحافيون أين يمكنهم العثور على المعلومات عندما يجرون بحثاً من أجل كتابة مقال. أما في صحافة البيانات، فأغلب الصحافيين لا يعلمون أي المعلومات ستكون متاحة.

معرفة تقنية بسيطة

تعتبر صحافة البيانات مجالاً تقنياً إلى حد كبير. ففي بعض الأحيانً عليك أن تستخرج البيانات، وفي أحيان أخرى عليك أن تقوم ببعض البرمجة لتوضيح بياناتك بصورة مرئية. ومن أجل إنجاز صحافة بيانات متميزة، ستحتاج إلى أمرين: الرؤية الصحفية لصحفي متمرس والمعرفة التقنية لخبير رقمي، وقد كان كلاهما متوافرا بكثرة أثناء ملتقى RegioHack.

هل هي أخبار؟

استخدم المشاركون مجموعة بيانات واحدة عموماً لاستكشاف الأخبار بدلاً من البحث عن الروابط المشتركة بين المصادر المختلفة. والسبب في ذلك هو أنك تحتاج إلى بعض المعرفة الإحصائية للتحقق من صحة الأخبار القادمة من صحافة البيانات.

ما هو الروتين؟

خلاصة كل ما ذكر أعلاه هو أنه لا يوجد روتين. فالمشاركون لديهم بعض المهارات ولكنهم لا يعلمون كيف ومتى يمكنهم استخدامها. وقد قارن أحد الصحافيين ذلك بعملية خبز كعكة قائلا: "لدينا جميعا المكونات: الدقيق والبيض والحليب وما إلى ذلك. وعلينا الآن أن نلقيها كلها في كيس ونرجها آملين أن تخرج منه كعكة." وحقا فإننا جميعنا لدينا المكونات ولكننا لا نعلم الوصفة اللازمة لإنجاز المهمة.

ما العمل الآن إذن؟ يمكن لتجاربنا الأولى في صحافة البيانات أن تساعد صحافيين آخرين أو مبرمجين يطمحون للدخول في نفس المجال، ونحن نعمل الآن لنصدر تقريراً بشأن ذلك.

كما أننا نفكر في كيفية مواصلة RegioHack في شكل ملتقى للمبرمجين. فقد وجدنا أنه ممتع وتعليمي ومثمر وبمثابة مقدمة رائعة لصحافة البيانات.

ولكن حتى تنجح صحافة البيانات، علينا أن ندمجها مع غرفة الأخبار. وينبغي على الصحافيين أن يفكروا من زاوية استخدام البيانات والأرقام، بالإضافة إلى الاقتباسات والتصريحات الصحفية واجتماعات المجالس وما إلى ذلك. وبإقامة ملتقى RegioHack، أثبتنا لجمهورنا أن صحافة البيانات ليست مجرد إثارة للجلبة. بل يمكننا كتابة مقالات أكثر إلماماً وتميزاً، لنقدم لقرائنا مقالات مختلفة سواء أكانت مطبوعة أو على الإنترنت.

—— جيري فيرمانين، موقع NU.nl

تتبع الأموال: صحافة البيانات والتعاون عبر الحدود

إن الصحافيين الاستقصائيين والمواطنين المهتمين بفضح الجريمة المنظمة والفساد اللذين يؤثران على حياة مليارات الناس في العالم، يصبح لديهم بمرور الأيام منافذ للحصول على المعلومات على نحو غير مسبوق. فالحكومات والمؤسسات المختلفة تتيح كماً هائلا من البيانات ويبدو أن المعلومات الضرورية تصبح في متناول الجميع بصورة متزايدة كل يوم. ولكن في الوقت ذاته، يفعل المسؤولون الحكوميون الفاسدون وجماعات الجريمة المنظمة كل ما بوسعهم لحجب المعلومات وإخفاء جرائمهم. كما أنهم يبذلون جهودا لإبقاء الناس في الظلام بينما يعقدون صفقاتهم القبيحة التي تتسبب في اضطرابات على جميع مستويات المجتمع وتؤدي إلى الصراعات أو المجاعات أو أنواع أخرى من الأزمات.

من الواجب إذن على الصحافيين الاستقصائيين أن يفضحوا تلك الاعتداءات وبذلك يعطلون الفساد والآليات الإجرامية.

هناك ثلاثة إرشادات أساسية قد تؤدي إلى صحافة جيدة ومتعمقة عند تقصّي أعمال الفساد والجرائم الكبرى حتى في أكثر الظروف قسوة.

فكر خارج بلدك

كثيراً ما يكون من الأسهل الحصول على المعلومات من الخارج بدلاً من الحصول عليها من داخل البلد التي يعمل بها الصحفي الاستقصائي. فالمعلومات التي يتم جمعها من الخارج عن طريق قواعد بيانات للمعلومات الأجنبية أو باستخدام قوانين المعلومات في بلاد أخرى قد تكون هي بالضبط ما تحتاجه لحل اللغز. فضلا عن ذلك، لا يستطيع المجرمون والمسؤولون الفاسدون حفظ أموالهم في نفس المكان الذي سرقوها منه. بل يحبذون إيداعها في بنوك أجنبية أو استثمارها في بلاد أخرى. ولأن الجريمة ظاهرة عالمية، فإن قواعد البيانات التي تساعد الصحافيين الاستقصائيين على تتبع الأموال حول العالم يمكن العثور عليها في عدة أماكن على الإنترنت. فعلى سبيل المثال، تتيح اللوحة الاستقصائية للصحافيين تتبع الأموال عبر الحدود (-http://www.investigativedash

.(/board.org /category /wwd



نموذج 2–12. اللوحة الاستقصائية (OCCRP)

الاستفادة من شبكات الصحافة الاستقصائية القائمة

تجمع الصحافيين الاستقصائيين حول العالم مؤسسات مثل 'مشروع إعداد التقارير حول الفساد والجريمة المنظمة' (.http://www.reportingproject net/) و'المنتدى الإفريقي للتقارير الاستقصائية' (.http://www.fairreporters org/) و'المراسلون العرب للصحافة الاستقصائية' (http://arij.net/) و'الشبكة العالمية للصحافة الاستقصائية' (http://www.gijn.org/). يمكن للصحافيين أيضاً أن يستفيدوا من منصات الصحافة الاحترافية مثل شبكة الصحافيين الدوليين IJNET، حيث يتم يوميا تبادل المعلومات المتعلقة بالصحافة العالمية. يعمل العديد من الصحافيين الذين تجمعهم الشبكات على قضايا متشابهة ويواجهون مواقف مماثلة لذا فمن المنطقي أن يتبادلوا المعلومات والوسائل. وترتبط بتلك الشبكات قوائم البريد الإلكتروني ومجموعات شبكات التواصل الاجتماعي، لذا

فمن السهل التواصل مع الزملاء الصحافيين وسؤالهم عن المعلومات أو طلب النصائح. يمكن أيضاً جمع أفكار الأخبار الاستقصائية من مثل تلك المنتديات وقوائم البريد الإلكتروني.

استفد من التكنولوجيا وتعاون مع المبرمجين

تساعد برامج الكمبيوتر الصحافيين على الوصول إلى المعلومات ومعالجتها. وتوجد أنواع عديدة من البرامج التي تساعد الصحفي الاستقصائي على فرز الكم الهائل من البيانات والبحث فيه ومحاولة فهمه، كما تساعد في العثور على الوثائق اللازمة لنشر الخبر. وهناك العديد من البرامج الجاهزة التي يمكن استخدامها كأداة لتحليل المعلومات أو جمعها أو تفسيرها، والأهم من ذلك، على الصحافيين الاستقصائيين أن يدركوا أن هناك أعداداً كبيرة من المبرمجين المستعدين لتقديم المساعدة عند الحاجة. هؤلاء المبرمجون يمكنهم الحصول على المعلومات والتعامل معها ويمكنهم تقديم مساعدة كبيرة فيما يتعلق بالجهد الاستقصائي. وبعض هؤلاء المبرمجين أعضاء في حركات عالمية للبيانات المفتوحة ويمكنهم أن يكونوا حلفاء فعالين في مكافحة الجريمة والفساد. ويمكنهم أيضاً أن يساعدوا الصحافيين على جمع المعلومات وتحليلها.

من الأمثلة الجيدة على واجهات التعامل بين المبرمجين والمواطنين موقع ScraperWiki (https://scraperwiki.com/) الذي يتيح للصحافيين طلب المساعدة من المبرمجين في استخراج البيانات من المواقع الإلكترونية. وتحتفظ اللوحة الاستقصائية بقائمة من الأدوات الجاهزة (http://bit.ly/dashboard-resources/) التي قد تساعد الصحافيين على جمع البيانات وصياغتها وتحليلها.

يمكن رؤية مدى الاستفادة من الإرشادات المذكورة أعلاه في الكثير من المواقف، وأحد الأمثلة الجيدة على ذلك عمل 'خديجة إسماعيلوفا'، وهي صحافية استقصائية آذرية ذات خبرة واسعة وتعمل في ظروف شديدة القسوة فيما يتعلق بالوصول إلى المعلومات. يتعين على الآنسة إسماعيلوفا تخطي العديد من العقبات يومياً حتى تتمكن من تقديم معلومات جيدة وموثوقة للشعب في أذربيجان. وفي يونيو 2011، قامت أثناء عملها كصحافية استقصائية في مكتب راديو أوروبا الحرة

في باكو، بنشر خبر يفيد (Radio Free Europe /Radio Liberty’s (RFE /RL بأن ابنتي الرئيس الأذري ‘إلهام علييف’ تُدِيران في سرية شركة اتصالات صاعدة تُدعى آذرفون Azerfon (http: / /bit.ly /rferl-azerfon) من خلال شركات خارجية تعمل في بَنَما. تفخر الشركة بأن لديها 1.7 مليون مشترك وتغطي 80٪ من مساحة البلاد وبأنها كانت حينها المزود الوحيد لخدمات 3G في أذربيجان. قضت إسماعيلوفا ثلاث سنوات في محاولة معرفة مُلاك الشركة ولكن الحكومة رفضت إعطاء أي معلومات عن حَمَلة الأسهم وكذبت عدة مرات بشأن ملكية الشركة، حتى إنها ادعت أن الشركة تملكها شركة سيمنز Siemens AG الألمانية، وهو الادعاء الذي نفته الشركة الألمانية بشكل قاطع. تمكنت الصحافية الآذرية من اكتشاف أن شركة آذرفون تملكها بضع شركات خاصة في بَنَما، وبدا أن تلك نهاية الطريق المسدود لتقريرها حتى تمت الاستعانة بمساعدة خارجية. في بدايات عام 2011، عندما علمت الآنسة إسماعيلوفا من خلال اللوحة الاستقصائية أنه بالإمكان تتبع الشركات في بَنَما، من خلال تطبيق (http: / /ohuiginn.net /panama/) طوره المبرمج والناشط دان أوهيجين، استطاعت أخيراً الكشف عن حقيقة أن ابنتي الرئيس ضالعتان في شركة الاتصالات من خلال شركات في بَنَما.

في الواقع، صنع أوهيجين أداة ساعدت الصحافيين من جميع أنحاء العالم على إعداد التقارير حول الفساد في بَنَما المعروفة بكونها ملاذاً خارجياً استغلها على نطاق واسع المسؤولون حول العالم كمكان آمن لإخفاء الأموال المنهوبة: بدءاً من المقربين إلى الرئيس المصري السابق حسني مبارك، وصولاً إلى المسؤولين الفاسدين في البلقان أو أمريكا اللاتينية. ما فعله المبرمج والناشط يسمّى استخراج البيانات من المواقع الإلكترونية (web scraping)، وهي وسيلة تتيح استخراج المعلومات وإعادة صياغتها حتى يتمكن الصحافيون الاستقصائيون من استخدامها. قام أوهيجين باستخراج البيانات من سجل شركات بَنَما (-http: / /www.reg istro-publico.gob.pa/) لأن هذا السجل - على الرغم من كونه مفتوحا - لا يسمح بالبحث سوى للصحافي الاستقصائي الذي يعرف اسم الشركة التجارية التي يبحث عنها. وقد أدى هذا إلى الحد من إمكانية التقصي لأن الصحافيين عادة ما يبحثون عن أسماء الأشخاص حتي يتتبعوا ممتلكاتهم. استخرج المبرمج البيانات

وأنشأ موقعاً إلكترونياً جديداً يتيح البحث بالاسم أيضاً. سمح الموقع الجديد للصحافيين الاستقصائيين في بلدان عديدة باصطياد المعلومات والبحث عن أسماء المسؤولين الحكوميين والبرلمانيين والتحقق مما إذا كانوا يملكون مؤسسات سرية في بَنَما (مثل عائلة الرئيس الآذري).

هناك مميزات أخرى لاستخدام الإرشادات المذكورة أعلاه، إلى جانب الوصول إلى المعلومات بشكل أفضل. إحداها تتعلق بتقليل الضرر وضمان حماية أفضل للصحافيين الاستقصائيين الذين يعملون في ظروف عدوانية. يرجع هذا إلى أن الصحافي لا يكون وحده عند العمل من خلال شبكة، بل يعمل مع زملاء في بلدان أخرى حتى يصعب على المجرمين أن يحددوا من الذي يعتقدون أنه وراء فضح جرائمهم. ونتيجة لذلك، فقد أصبح انتقام الحكومات أو المسؤولين الفاسدين أصعب.

هناك أمر آخر يجب وضعه في الاعتبار وهو أن المعلومات التي لا تبدو ذات قيمة في منطقة جغرافية ما قد تكون ذات أهمية قصوى في منطقة أخرى. قد يؤدي تبادل المعلومات من خلال شبكة استقصائية إلى نشر أخبار هامة للغاية. فعلى سبيل المثال، لن يكون خبر القبض على روماني في كولومبيا لحيازته على كيلوغرام من المخدرات مهماً في بوغوتا، ولكنه قد يكون في غاية الأهمية للرومانيين إذا تمكن صحفي روماني من معرفة أن الشخص الذي تم القبض عليه لحيازته على مخدرات موظف حكومي في بوخارست.

وعليه يمكن القول إن الصحافة الاستقصائية الفعّالة هي نتيجة للتعاون بين الصحافيين الاستقصائيين والمبرمجين وغيرهم من الذين يريدون استخدام البيانات من أجل خلق مجتمع عالمي أكثر نزاهة وعدلاً ومساواة.

— بول رادو، مشروع إعداد التقارير حول الفساد والجريمة المنظمة

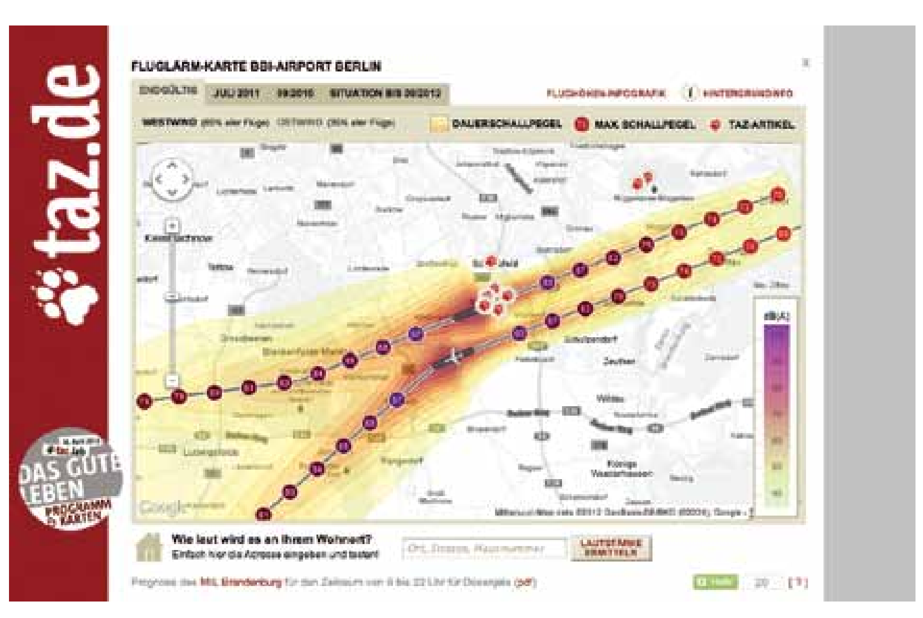
تأتي قصصنا الإخبارية في صيغة كود

تأسست "مدينة البيانات الحرة" OpenDataCity (http://www.opendatacity.de/) في أواخر عام 2010. ولم يكن هناك شيء يُذكر يمكنك أن تطلق عليه اسم صحافة البيانات في ألمانيا آنذاك.

لماذا فعلنا ذلك؟ كثيراً ما سمعنا العاملين في الصحف والإذاعات يقولون: "لا، لسنا مستعدين لبدء وحدة خاصة لصحافة البيانات في غرفة الأخبار. ولكن يسرنا أن نستعين بمصادر خارجية."

نحن، فيما نعلم، الشركة الوحيدة المتخصصة حصرياً في صحافة البيانات في ألمانيا. نحن الآن ثلاثة أشخاص: اثنان لديهما خلفية صحافية والثالث لديه فهم عميق للترميز بالكود والتوضيح المرئي، ونعمل مع مجموعة من المبرمجين والمصممين والصحافيين المستقلين.

خلال الاثني عشر شهراً الماضية، تولينا أربعة مشاريع لصحافة البيانات في الصحف، وقدمنا دورات تدريبية وخدمات استشارية للإعلاميين والعلماء وكليات الصحافة. أول تطبيق أنشأناه كان يدعى TAZ، وهو أداة تفاعلية تتعلق بالضجة المحيطة بالمطار المبني حديثا في برلين (http://bit.ly/taz-airport-noise). كان مشروعنا المميز التالي تطبيقاً لحفظ معلومات استخدام الهاتف المحمول (http://bit.ly/zeit-telephone) لسياسي ألماني مع صحيفة دي تسايت الإلكترونية. وبسبب هذا التطبيق فزنا بجائزة غريمي أونلاين Grimme



نموذج 2–13. خريطة الضجة المحيطة بالمطار في برلين. (Taz.de)

وجائزة Lead في ألمانيا (http://bit.ly/grimme-award) Online Award وجائزة الصحافة الإلكترونية من جمعية الصحافة الإلكترونية (http://bit.ly/online-news-award) في الولايات المتحدة. وبينما تستمر كتابة هذا الكتاب، لدينا العديد من المشاريع قيد التنفيذ، وهذه تتراوح بين رسوم بيانية تفاعلية بسيطة وتصميم نوع من البرمجيات الوسيطة لصحافة البيانات وتطويرها.

بالطبع يساعد الفوز بالجوائز على بناء اسم وسمعة. ولكن حين نخاطب الناشرين الذين عليهم الموافقة على المشاريع، لا تعتمد حجتنا للاستثمار في صحافة البيانات على الفوز بالجوائز. بل تعتمد على جذب الانتباه لفترة زمنية طويلة وعلى نحو مستدام، وإحداث تراكم، وليس الاكتفاء بالسبق الصحفي الذي يُنسى في غضون أيام معدودة.

فيما يلي ثلاث حجج استخدمناها لتشجيع الناشرين على تولي مشاريع طويلة الأمد:

مشاريع البيانات لا تتقادم بمرور الوقت

يمكن إضافة مواد جديدة إلى تطبيقات صحافة البيانات، اعتماداً على تصميمها. وهي ليست للمستخدمين فقط، بل يمكن استخدامها داخلياً من أجل إعداد التقارير والتحليلات. وإذا كنت تخشى من أن يؤدي هذا الأمر إلى استفادة منافسيك من استثمارك، يمكنك إبقاء بعض الخصائص أو بعض البيانات للاستخدام الداخلي فقط.

يمكنك الإضافة إلى عملك السابق

عندما تتولى مشروع بيانات، عادة ما ستكتب أجزاء من رموز التشفير التي يمكن إعادة استخدامها أو تحديثها. وقد يستغرق المشروع التالي نصف الوقت، لأنك تعلم بشكل أفضل ما يجب فعله (وما لا يجب فعله) ولديك قطع وأجزاء يمكنك الإضافة إليها.

صحافة البيانات تتكفل بتكاليفها

تعد المشاريع المبنية على البيانات أقل تكلفة من حملات التسويق التقليدية.

وعادة ما تستثمر مصادر الأخبار على الإنترنت في أمور مثل تحديث محرك البحث (Search Engine Optimization (SEO وتسويق محرك البحث Search (Engine Marketing (SEM. عادة ما يؤدي مشروع البيانات المُنَفذ إلى إثارة الكثير من الاهتمام والزيارات الإلكترونية، وقد ينتشر سريعاً. وفي العادة، هذا يكلف الناشرين مبلغاً أقل مما إذا حاولوا إيجاد نفس القدر من الانتباه وجذب الاهتمام عبر الزيارات والروابط عن طريق تسويق محرك البحث (SEM).

لا يختلف عملنا كثيراً عن الوكالات الإعلامية الأخرى: أي توفير التطبيقات أو الخدمات لمصادر الأخبار. ولكننا قد نختلف في أننا نعتبر أنفسنا صحافيين في المقام الأول. نحن نرى أن ما نقدمه من منتج هو مقالات أو أخبار، حتى وإن لم تكن في صيغة كلمات وصور أو صوت أو فيديو، بل في صيغة رموز او شفرات كودية. حين نتحدث عن صحافة البيانات، علينا أن نتحدث عن التكنولوجيا وبرامج الكمبيوتر والأجهزة وكيف يمكننا استخدامها لسرد الأخبار.

على سبيل المثال: انتهينا لتونا من العمل على تطبيق يحصل على بيانات الزمن الحقيقي realtime data من خلال استخراج بيانات من الموقع الإلكتروني لهيئة سكك الحديد الألمانية، مما يسمح لنا بإعداد 'مراقب القطارات' Train Monitor لموقع SüddeutscheZeitung (http://zugmonitor.sueddeutsche.de/)، الذي يوضح تأخر قطارات المسافات البعيدة في صيغة الزمن الحقيقي. يتم تحديث بيانات التطبيق كل دقيقة، كما نزوده بخطوط اتصال لبرمجة التطبيقات (API). بدأنا في ذلك منذ عدة أشهر وقد جمعنا حتى الآن مجموعة بيانات هائلة تنمو باستمرار، وتتراكم لتكوّن مئات الآلاف من صفوف البيانات. يتيح المشروع للمستخدم أن يستكشف البيانات بزمنها الحقيقي وأن يبحث في أرشيف الأشهر الماضية. وفي النهاية سيتم تحديد الخبر الذي نسرده حسب سلوك الأفراد المستخدمين وما يقومون به.

في الصحافة التقليدية، وبسبب الطابع الخطي للإعلام المكتوب أو المُذاع، علينا أن نفكر بالبداية والنهاية وحبكة الخبر وطول التقرير وزاويته. أما بالنسبة لصحافة البيانات فالأمر يختلف. نعم هناك بداية، ويأتي الناس إلى الموقع الإلكتروني ويكوّنون انطباعا أولياً عن الواجهة. ولكنهم حينها يصبحون وحدهم،

وقد يبقون لدقيقة أو لنصف ساعة.

يقتضي عملنا كصحافيين متخصصين في البيانات أن نوفر إطار العمل أو الظروف المناسبة لتحقيق ذلك. وبالإضافة إلى الأجزاء المتعلقة بالرموز ومكونات إدارة البيانات، علينا أن نفكر في طرق ذكية لتصميم التجربة. فتجربة المستخدم (User Experience (UX مشتقة أساساً من وصلة المستخدم (للرسوم) (-Graph ical) User Interface (GUI). وفي النهاية هذا هو الجزء الذي سيؤدي إلى نجاح أو فشل المشروع. قد تكون لديك في الخلفية أفضل رموز الكود للتعامل مع مجموعة بيانات مثيرة، ولكن إذا كانت الواجهة سيئة، فلن يكترث أحد بها.

لا يزال هناك الكثير لنتعلمه ونختبره. ولكن لحسن الحظ توجد صناعة الألعاب التي تقدم ابتكارات تتعلق بالسرد الرقمي والبيئة الرقمية والواجهات منذ عدة عقود. لذا فإننا حين نطور تطبيقات صحافة البيانات علينا أن نراقب عن كثب كيفية عمل تصميمات الألعاب وكيف يتم السرد فيها. لماذا تكون ألعاب عادية مثل لعبة تيتريس Tetris ممتعة؟ ولماذا تكون ألعاب العالم المفتوح مثل سرقة السيارات ((Grand Auto Theft أو سكايريم Skyrim جاذبة؟

نرى أن صحافة البيانات باقية، وفي خلال بضع سنوات سيتم دمج مهام صحافة البيانات في غرفة الأخبار بشكل طبيعي، لأنه سيكون على المواقع الإلكترونية أن تتغير. فكم البيانات المتاحة للجمهور سيزداد باستمرار. ولكن لحسن الحظ ستواصل التكنولوجيا الجديدة السماح لنا بالعثور على سبل جديدة للسرد. ستعتمد بعض الأخبار على البيانات وسيكتسب العديد من التطبيقات والخدمات طابعاً صحفياً. السؤال المثير للاهتمام هو: ما هي الاستراتيجية التي ستتبناها غرف الأخبار لتبني هذه العملية؟ هل سيشكلون فرقاً من صحافيي البيانات لدمجهم في غرف الأخبار؟ هل ستكون هناك أقسام للبحث والتطوير أشبه بشركات ناشئة داخلية؟ أم ستتم الاستعانة بشركات متخصصة لتتولى جزءاً من العمل؟ ما زلنا في البداية ولن نعرف الإجابة إلا بمرور الوقت.

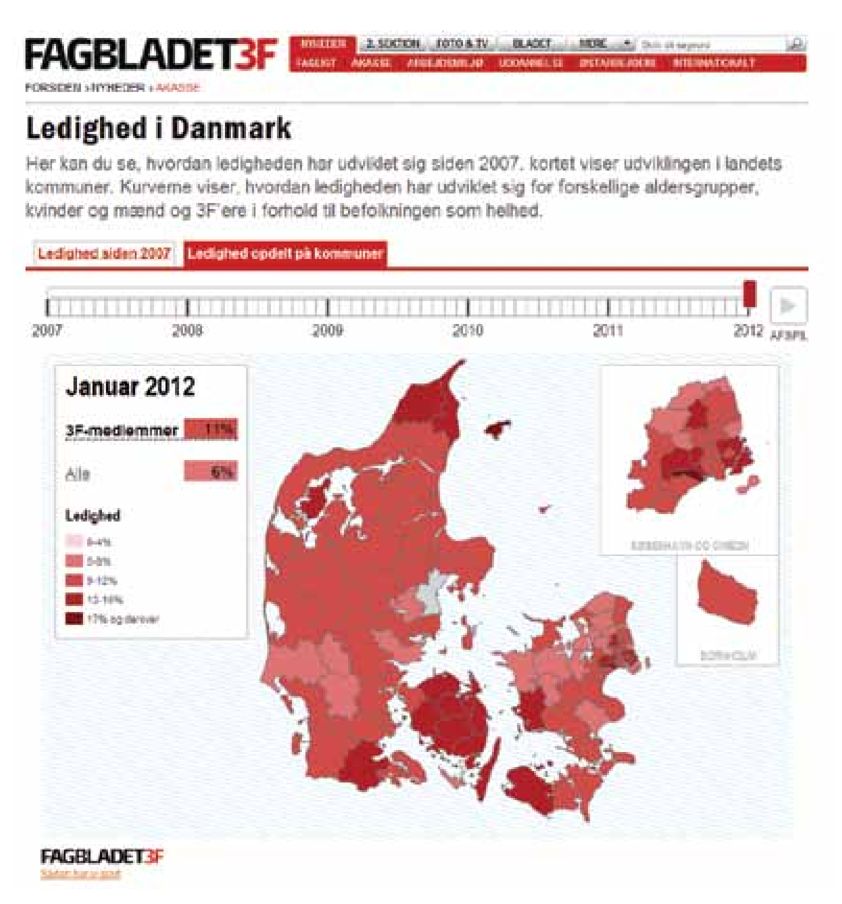
—— لورينز ماتزات، OpenDataCity

كاس وملفاد: تقديم محتوى أولي يخدم الإعلام الموجه

يُعتبر الإعلام الموجّه قطاعاً صاعداً عادة ما يتغاضى عنه واضعو النظريات، رغم أن لديه إمكانية إحداث تأثير هائل، سواء عن طريق شبكة الإنترنت أو تزويد وسائل الإعلام بالمحتوى المطلوب. ويمكن تعريفه على أنه إعلام (غالباً على الإنترنت) تسيطر عليه منظمات أو مؤسسات من أصحاب التأثير (في أي نشاط تجاري فإن هؤلاء هم الزبائن، العاملون، الموردون، المجتمع المحلي، المساهمون، الخ)، ويتم استغلاله لدعم مصالح أو مجتمعات معينة. عادة ما تقوم المنظمات غير الحكومية بإنشاء مثل تلك الوسائل الإعلامية، وكذا تفعل ذلك أيضا مجموعات المستهلكين والجمعيات المهنية والنقابات العمالية وما شابه. وأهم ما يحد من قدرته على التأثير على الرأي العام أو أصحاب الشأن الآخرين هو افتقاده إلى القدرة على العمل لاكتشاف معلومات هامة، وهذه القدرة تقل حتى عن وسائل الإعلام التي لجأت إلى تقليص العمالة. تعتبر مؤسسة كاس وملفاد Kaas&Mulvad، وهي مؤسسة دنماركية ربحية، من أوائل شركات الإعلام الاستقصائي التي توفر الخبراء لتلك المصادر الإعلامية التابعة لأصحاب التأثير.

تأسست الشركة عام 2007 من وحي 'المعهد الدنماركي للصحافة بمساعدة الكمبيوتر' Danish Institute for Computer Assisted Reporting (Dicar)، وهو معهد لاربحي، كان يبيع تقارير استقصائية للإعلام ويدرّب الصحافيين على استخدام صحافة البيانات. كان مؤسسا المعهد وهما طوميكاس ونيل ملفاد صحافيين في مجال الأخبار فيما سبق. تقدم شركتهما الجديدة ما يطلقان عليه 'البيانات والرؤية الصحفية' (حيث المحتوى نصف مكتمل ويتطلب المزيد من التنقيح وإعادة الكتابة) إلى الإعلام الموجّه، التابع لأصحاب التأثير، الذي يقوم بدوره بإكمال المحتوى ليصبح نشرات إخبارية أو تقارير ويوزعها من خلال وسائل الإعلام ومصادرها الخاصة بها (كالمواقع الإلكترونية). ومن زبائن الشركة المباشرين مؤسسات حكومية وشركات علاقات عامة ونقابات عمالية ومنظمات غير حكومية مثل سجل شفافية الاتحاد الأوروبي EU Transparency والصندوق العالمي للحياة البرية World Wildlife Fund. يتضمن عمل المنظمة غير الحكومية رصد دعم المزارع ومصائد الأسماك، وأخبار منتظمة عن نشاطات

جماعات الضغط في الاتحاد الأوروبي من المواقع الإلكترونية ذات الصلة. ومن الزبائن غير المباشرين المؤسسات التي تمول مشاريع المنظمات غير الحكومية. كما تعمل الشركة أيضاً مع صناعة الأخبار؛ فعلى سبيل المثال قامت صحيفة صفراء بشراء خدمات مراقبة المشاهير الخاصة بالشركة.



نموذج 2–14. شركات الإعلام الموجه (Fagbladet3F)

من ضمن مشاريع صحافة البيانات لدى كاس وملفاد:

خريطة البطالة لنقابة 3F (http://bit.ly/3F-unemployment)

توضيح مرئي للبيانات مع أرقام هامة عن البطالة في الدنمارك لنقابة 3F، وهي نقابة العمال غير المهرة في الدنمارك.

ظروف المعيشة لنقابة 3F (http://bit.ly/3F-living)

مشروع آخر لنقابة 3F يوضح اختلاف ظروف المعيشة مع اختلاف المناطق في الدنمارك. تظهر الخريطة 24 مؤشراً مختلفاً لظروف المعيشة.

الديون لـ "Ugebrevet A4"(http://bit.ly/3F-debt-index)

مشروع يقوم بحساب 'مؤشر الديون' ويوضح مرئياً الاختلافات في الاقتصاد الخاص.

المنشآت الخطرة في الدنمارك (http://bit.ly/3F-dangerous-facilities)

مشروع يرسم خريطة ويحلل مدى قرب المنشآت الخطرة من رياض الأطفال ودور الحضانة الأخرى، كان المشروع لمجلة "Børn&Unge" التي تنشرها BUPL وهي النقابة الدنماركية لمعلمي الأطفال والشباب.

بيانات مسؤولية الشركات لمؤسسة فيستاس Vestas (http://data.vestas.com)

توضيح مرئي لخمسة جوانب من بيانات مسؤولية الشركات لمؤسسة فيستاس الدنماركية لتوربينات الرياح، ومعه نص تلقائي. يتم تحديث كل هذا تلقائيا كل ثلاثة أشهر في 400 صفحة إلكترونية بدءاً من النطاق العالمي وصولاً إلى وحدات الإنتاج الفردية.

خريطة الأسماء لشركة اكسبريان Experian (http://xpoint.experian.dk/navnekort)

اكتب اسم عائلتك وانظر إلى توزيع هذا الاسم على مناطق جغرافية مختلفة في الدنمارك.

خريطة "smily" الوجوه المبتسمة لصحيفة EkstraBladet (http://ekstrabladet.dk/kup/fodevarer/)

نقوم يوميا باستخراج بيانات كل عمليات التفتيش عن الطعام الفاسد وننشر

أحدثها في خريطة للصحيفة الدنماركية EkstraBladet (انظر إلى منتصف صفحة الويب لتجد الخريطة).

ليس كاس وملفاد أول صحافيين يعملان مع الإعلام الموجه. فمنظمة السلام الأخضر Greenpeace مثلاً تستعين بصحافيين بانتظام ليعدوا تقارير لها. ولكننا لم نسمع عن أي شركة أخرى تعتمد على البيانات فيما تقدمه للإعلام الموجه لأصحاب التأثير. فمن العادي أكثر أن يعمل الصحافيون مع المنظمات غير الحكومية كمعدّي تقارير ومحررين وكُتّاب. من الواضح أن وسائل الإعلام بمساعدة الكمبيوتر حالياً تركز على البحث والاكتشاف (مثل ويكيليكس)؛ وهنا يبتكر كاس وملفاد مرة أخرى عن طريق التركيز على تحليل البيانات. ولكن منهجهما لا يتطلب مهارات البرمجة فحسب، بل يتطلب أيضاً فهم نوع المعلومات التي يمكن من خلالها صناعة خبر له تأثيره. وعليه يمكننا القول إنه إن أراد أحد أن يقلد خدماتهما فعليه على الأرجح أن يكتسب مجموعتي المهارات من خلال الشراكة مع طرف آخر، لأن الفرد نادراً ما يملك كليهما.

العمليات: تكنولوجيا معلومات مبتكرة بالإضافة إلى التحليل

تتولى الشركة حوالي 100 مشروع سنوياً، تتراوح مدتها ما بين بضع ساعات إلى بضعة أشهر. كما أنها تستثمر دائماً في مشاريع من شأنها أن توسع قدراتها وما يمكنها تقديمه. وكانت خدمة تعقّب المشاهير إحدى تلك التجارب. هناك تجربة أخرى كانت تتعلق باستخراج البيانات من الإنترنت للحصول على أخبار تتعلق باستعادة البنوك ملكية المنازل المرهونة بسبب عدم سداد الأقساط، بالإضافة إلى رسم خرائط للأحداث. يقول الشريكان إن معيارهما الأول لاختيار المشاريع هو ما إذا ما كانا يستمتعان بالعمل عليه ويتعلمان منه شيئا جديدا؛ فالأسواق يُسعى إليها حين يتم تقديم خدمة جديدة. وقد أوضحا أنه في مجال صناعة الأخبار يصعب ابتكار أساليب عمل ونشاطات جديدة.

يقول ملفاد معلقاً:

"ليس لدينا محررون أو رؤساء عمل يحددون أي المشاريع يمكننا تولّيها أو أي برامج أو أجهزة يمكننا شراؤها. فنحن نستطيع شراء الأدوات وفقاً لما يحتاجه

المشروع كأفضل الحلول لاستخراج البيانات والتنقيب عنها. نهدف إلى التقدم في تلك المجالات ونحاول جذب زبائن على استعداد للدفع، أو إذا كان المشروع مسلياً نقوم به لقاء مبلغ أقل."

القيمة الناتجة: العائدات المالية والعلامات التجارية للشركات أو الأشخاص

بلغ إجمالي الربح عام 2009 حوالي مليونين ونصف مليون كرونة دنماركية أو ما يعادل 336,000 يورو. وتركز الشركة على إبراز سمعة الشريكين باعتبارهما صحافيين متميزين، مما يحافظ على الطلب على خدماتهما للتدريس وإلقاء المحاضرات. كما أن ظهورهما في المناسبات العامة يدعم العلامة التجارية للشركة.

الأفكار الرئيسية حول هذا المثال

- إن أزمة تدهور قدرات أجهزة الأخبار هي أيضاً أزمة سببها عدم استغلال القدرات والاستفادة من الكفاءات. فقد اضطر كاس وملفاد لترك صناعة الأخبار ليتسنّى لهما القيام بعمل يقدّران قيمته ويدركان إمكانية أن ينتج عنه عائد. ولم يكن هناك ما يمنع أي مؤسسة إخبارية من اقتناص هذه القيمة.
- في بضع أسواق على الأقل، توجد سوق مربحة للمحتوى 'شبه الكامل' الذي من شأنه أن يخدم مصالح المجموعات صاحبة التأثير.
- ولكن هذه الفرصة تطرح قضية تتعلق بمدى سيطرة الصحافيين على تقديم أعمالهم واستخدامها من قبل أطراف أخرى. نود أن نذكّر بأن تلك القضية موجودة بالفعل داخل صناعة الأخبار (حيث يستطيع المحررون فرض تغييرات على ما ينتجه الصحافي)، وهي موجودة داخل صناعات إعلامية أخرى (مثل صناعة الأفلام حيث تكثر الصراعات بين المخرجين والمنتجين حول 'النسخة النهائية'). وهذه الظاهرة لا تشكل خطراً أخلاقياً معيناً على الإعلام الموجه، ولكنها لن تختفي أيضاً. لذلك يجب توجيه المزيد من الانتباه لأخلاقيات هذا الواقع المتنامي ومتطلبات السوق.
- ومن جهة العائدات، لا يكفي منتج واحد أو خدمة واحدة. وعلى شركات مراقبة البيانات الناجحة أن تعمل بشكل أفضل لتتبنى منهجاً شاملاً ومتنوعا حيث تقوم

أيضا بأعمال الاستشارات والتدريس وإلقاء المحاضرات والخدمات الأخرى، لتوفر عائدات إضافية وتدعم العلامة التجارية.

—— مقتطفات منقحة من كتاب 'تكنولوجيا الأخبار المشوشة: الإعلام الموجه ومستقبل نماذج أعمال شركات المراقبة الصحافية.' مارك لي هنتر ولوك ن. فان واسنهوف. طبعة INSEAD، 2010.

نماذج أنشطة تجارية ناجحة لصحافة البيانات

وسط كل الاهتمام والأمل فيما يخص الصحافة المبنية على البيانات، يوجد سؤال يثير فضول غرف الأخبار بشكل دائم: ما هي نماذج الأعمال؟

وبينما علينا توخي الحذر عند وضع التكهنات، فالنظر إلى التاريخ الحديث والوضع الحالي لصناعة الإعلام يمكن أن يساعدنا على الفهم. توجد اليوم عدة مؤسسات إخبارية حققت نجاحات مادية ملموسة بسبب تبنّيها أساليب جديدة لتناول القضايا.

وقد تبدو مصطلحات مثل 'صحافة البيانات' و'علم البيانات' التي راجت مؤخرا كأنها تصف أمراً جديداً، وهذا ليس دقيقا. فهذه التسميات الجديدة هي مجرد طرق لوصف تغيير ازداد قوة عبر عقود.

يجهل الكثير من الصحافيين حجم العائدات التي أتت من جمع البيانات وتحليلها وتوضيحها مرئياً، وهذه هي تجارة تنقية المعلومات وتهذيبها. ومع ظهور أدوات وتكنولوجيا البيانات، تزداد إمكانية تسليط الضوء على قضايا شديدة التعقيد سواء أكانت تدور حول التمويل الدولي أم الديون أم الإحصاء السكاني أم التعليم، الخ. يصف مصطلح 'استخبارات الأعمال' "business intelligence" مجموعة من المفاهيم تتعلق بتكنولوجيا المعلومات التي تهدف إلى توفير رؤية واضحة لما يحدث في المؤسسات التجارية. فالشركات الكبرى والمربحة في عصرنا هذا – ومن بينها ماكدونالدز أو زارا أو اتش آند إم – تعتمد على تتبع البيانات لتحقيق الأرباح، والأمر ناجح جداً مع تلك الشركات.

وما يتغير الآن هو أن الأدوات التي طورت من أجل هؤلاء أصبحت في متناول مجالات أخرى من بينها الإعلام، وهناك صحافيون يفهمونها جيدا. فشركة تابلو

Tableau على سبيل المثال توفر مجموعة من أدوات التوضيح المرئي، أو حركة 'البيانات الكبيرة' Big Data حيث تقوم شركات التكنولوجيا باستخدام باقات برامج (عادة ما تكون من مصدر مفتوح open source) للتنقيب في أكوام البيانات واستخراج الأفكار والمؤشرات خلال أجزاء من الثانية.

يمكن تطبيق تلك التكنولوجيا الآن على الصحافة. وتواصل الفرق في صحيفتي الغارديان والنيويورك تايمز تخطي الحدود في هذا المجال الصاعد. وما نراه الآن هو مجرد البداية.

ولكن كيف يمكن لهذا أن يحقق الأرباح للصحافة؟ إن السوق العالمية الكبيرة التي تتفتح الآن تتعلق بتحويل البيانات المتاحة للجمهور إلى شيء يمكننا معالجته: أي جعل البيانات مرئية وذات مدلولات إنسانية. نريد أن نتمكن من التعرف على الأرقام الكبيرة التي نسمعها يوميا في الأخبار وما تعنيه تلك الملايين بالنسبة لكل واحد منا.

هناك عدد من الشركات الإعلامية المربحة التي تعتمد على البيانات والتي قامت بكل بساطة بتطبيق هذا المبدأ قبل الآخرين. وهي تتمتع بمعدلات نمو سليمة وأحياناً تحقق أرباحاً تثير الإعجاب. أحد الأمثلة شركة بلومبرج التي تدير حوالي 300,000 نظام كمبيوتر وتقدم البيانات المالية لمستخدميها، وهذه أداة قوية إذا كنت تعمل في مجال المال والبورصات. يأتي كل نظام مع لوحة مفاتيح مشفرة بالألوان وحتى 30,000 خيار للبحث والمقارنة والتحليل، ولمساعدتك على تحديد الخطوة التالية. تحقق هذه التجارة الأساسية أرباحاً تقدر بحوالي 6.3 مليار دولار سنويا – على الأقل هذا ما قدره مقال عام 2008 في صحيفة نيويورك تايمز (http://nyti.ms/IQcRgY). ونتيجة لذلك، بدأت شركة بلومبرج بتوظيف صحافيين من كل مكان واشترت مجلة Business Week القديرة ولكن غير المربحة.

مثال آخر: مجموعة شركات الإعلام الكندية المعروفة اليوم باسم طومسون رويترز Thomson Reuters. فقد بدأوا بصحيفة واحدة واشتروا عدداً من الأسماء المعروفة في المملكة المتحدة ثم قرروا منذ 20 سنة أن يتركوا مجال الصحف. وبدلاً من ذلك، أسسوا نموهم على خدمات المعلومات هادفين إلى توفير وجهات نظر أعمق للزبائن من شتى المجالات. إذا كنت قلقاً إزاء كيفية الربح من المعلومات

المتخصصة، فننصحك فقط بقراءة تاريخ تلك الشركة في موقع ويكيبيديا (http://
en.wikipedia.org/wiki/The_Thomson_Corporation).

وانظر أيضاً إلى مجلة الإيكونوميست، فقد رسخت المجلة في جانبها الإعلامي علامة تجارية ممتازة ومؤثرة. وفي الوقت ذاته، أصبحت 'وحدة معلومات إيكونوميست' "Economist Intelligence Unit" أشبه بشركة استشارية تعد تقارير عن التوجهات والتنبؤات ذات الصلة لأي دولة في العالم تقريباً. وهُم يوظفون مئات الصحافيين ويزعمون أنهم يخدمون حوالي المليون ونصف المليون من الزبائن حول العالم.

وتوجد عدة خدمات ملائمة مبنية على البيانات ويمكن أن تكون مصدر إلهام: منها eMarketer في الولايات المتحدة التي توفر مقارنات ورسوم بيانية ونصائح للمهتمين بالتسويق عبر الإنترنت، وStiftungWarentest في ألمانيا، وهي مؤسسة تبحث في جودة المنتجات والخدمات، وStatista أيضاً في ألمانيا، وهي شركة ناشئة تساعد على التوضيح المرئي للمعلومات المتاحة للجمهور.

هناك حالياً موجة حول العالم من الشركات الناشئة في هذا القطاع، وهي بطبيعة الحال تغطي مجموعة واسعة من المجالات – على سبيل المثال Timetric التي تهدف إلى 'إعادة تشكيل أبحاث التجارة'، وOpenCorporates وKasabi وInfochimps وData Market. يمكن القول إن العديد من تلك الشركات هي مجرد تجارب، ولكن يمكن اعتبارها جميعا علامات هامة على طريق التغيير.

وهناك الإعلام العام الذي يعتبر العملاق النائم فيما يخص الصحافة المبنية على البيانات. في ألمانيا يجني هذا القطاع 7.2 مليار يورو سنويا. فالصحافة منتج مميز: إذا تمت بشكل جيد، فلن تقتصر على جني المال بل إنها تؤدي دوراً هاماً في المجتمع أيضاً. وبمجرد توضيح فكرة أن صحافة البيانات يمكنها توفير أفكار ورؤية أفضل بشكل أسهل ويُعتمد عليها، يمكن استخدام بعض هذه الأموال لخلق وظائف جديدة في غرف الأخبار.

فيما يخص صحافة البيانات، لا يتعلق الأمر بكون المؤسسة هي السباقة والأولى في نشر الخبر، ولكن أيضاً بكونها مصدراً موثوقاً للمعلومات. ففي هذا العالم المتعدد القنوات، يمكن أن يتوفر الاهتمام بكثرة، ولكن الثقة تزداد ندرة.

يمكن للصحافيين أن يساعدوا في المقارنة والتركيب والتقديم لموارد المعلومات المتنوعة والصعبة عادة، بطريقة تقدم لجمهورهم رؤية حقيقة للقضايا المعقدة. وبدلاً من مجرد إعادة صياغة التصريحات الصحفية وإعادة سرد الأخبار التي سمعوها من مصدر آخر، يمكن لصحافيي البيانات أن يقدموا للقراء وجهة نظر واضحة ومفهومة ويُفضّل أن تكون قابلة للتخصيص مع رسوم بيانية تفاعلية ودخول مباشر للمصادر الرئيسية. وهذا ليس أمراً عادياً، بل قيّما بالتأكيد.

إذن فما هو المنهج الأفضل لصحافيي البيانات الطامحين ليستكشفوا هذا المجال ويقنعوا إداراتهم بدعم المشاريع المبتكرة؟

الخطوة الأولى هي البحث عن فرص مباشرة وقريبة الأهداف. فعلى سبيل المثال، قد تكون لديك بالفعل مجموعات من النصوص المنظمة والبيانات التي يمكنك استخدامها. ومن الأمثلة الرئيسية على ذلك 'قاعدة بيانات جرائم القتل' في صحيفة لوس أنجلس تايمز. وفي هذه الحالة تكون البيانات والتوضيحات المرئية هي الجوهر وليست فكرة ملحقة. يقوم المحررون بجمع كل حالات الجرائم الموجودة وحينها فقط يكتبون مقالات مبنية عليها. ومع مرور الوقت، يزداد حجم تلك المجموعات وتصبح أفضل وأعمق وأكثر قيمة.

قد لا ينجح هذا من المرة الأولى، ولكنه سينجح مع مرور الموقت. ومن المؤشرات الواعدة هنا أن صحيفتي تكساس تريبيون وProPublica، اللتين يمكن القول بأنهما شركتان إعلاميتان في مرحلة ما وراء الطباعة، صرحا بأن تمويل مؤسستيهما الصحفيتين غير الربحيتين تجاوز أهدافهما في زمن أقل مما كان مخططا له.

إن المهارة في كل ما يتعلق بالبيانات – سواء أكانت على أساس عام أو متخصص بالتركيز على جانب واحد من سلسلة البيانات – توفر منظوراً قيّماً لمن يؤمن بالصحافة. وقد صرح ناشر معروف في ألمانيا في مقابلة بأن "هناك مجموعة جديدة تطلق على نفسها اسم صحافيي البيانات وهم ليسوا على استعداد للعمل بأجر زهيد بعد الآن."

—— ميركو لورينز، دويتشه فيله

الفصل الثالث

دراسات لحالات نموذجية



في هذا الجزء سنلقي نظرة عميقة، ونرصد ما خلف الأحداث حول العديد من مشروعات صحافة البيانات، بدءاً بالتطبيقات التي يتم تطويرها خلال يوم واحد وحتى التحقيقات التي تمتد إلى نحو تسعة أشهر. سنطّلع على كيفية استخدام مصادر البيانات لتعزيز وتحسين التغطية لكافة الأخبار بما في ذلك أخبار الانتخابات والإنفاق العام وحتى أعمال الشغب والفساد وأداء المدارس وأسعار المياه. وكما نتعلم من المؤسسات الإعلامية الكبيرة، مثل بي. بي. سي والصحف مثل شيكاغو تريبيون والغارديان والفاينانشال تايمز وهيلسنغين سانومات ولاناسيون وول ستريت جورنال وصحيفة دي تسايت الإلكترونية، سنتعلم أيضاً من أداء المبادرات الصغيرة مثل كاليفورنيا ووتش وملتقى المبرمجين والصحافيين (Hacks/Hackers) في بوينوس آيريس وبروبابليكا (ProPublica) ومجموعة محلية من الصحافيين المواطنين في البرازيل والمعروفين بـ "أصدقاء جانواريا".

عدم تكافؤ الفرص

استخدم مشروع «عدم تكافؤ الفرص» (http://projects.propublica.org/schools) بيانات من وزارة التعليم الأمريكية لم يسبق نشرها من قبل وتتعلق بالحقوق المدنية، حيث أوضحت تلك البيانات أن بعض الولايات مثل فلوريدا قد حققت المساواة وقدمت للطلاب الأغنياء والفقراء على حد سواء فرصاً متساوية تقريباً للالتحاق بالدراسات عالية المستوى. أما الولايات الأخرى مثل كانساس وميريلاند وأوكلاهوما، فتوفر قدراً أقل من فرص التعليم في المناطق التي تسكنها العائلات الفقيرة.

نموذج 3–1. مشروع "عدم تكافؤ الفرص" (بروبابليكا)

شملت البيانات كافة المدارس الحكومية في منطقة بها 3,000 طالب أو أكثر حيث تم تمثيل أكثر من ثلاثة أرباع تلاميذ المدارس الحكومية. وقد حصل صحفي من غرفة الأخبار لدينا على البيانات، بينما تولى مدير الصحافة بمساعدة

الكومبيوتر تنقية البيانات بدقة.

استغرق تنفيذ المشروع حوالى ثلاثة أشهر حيث عمل ستة أشخاص على تحرير الخبر وبناء تطبيق الأخبار: اثنان من المحررين وصحفي واختصاصي "صحافة بمساعدة الكومبيوتر" ((CAR واثنان من المطوّرين. لم يكن معظمنا يعمل على هذا المشروع بشكل حصري خلال تلك الفترة.

تطلب هذا المشروع استخدام مجمل مهاراتنا: أي المعرفة التامة بمجال العمل مع فهم أفضل ممارسات استخدام البيانات، ومهارات في التصميم والترميز (كتابة الكود) وما إلى ذلك. والأهم من ذلك أنه تطلّب القدرة على العثور على الخبر داخل البيانات. كما تطلب الأمر القيام بالتحرير، ليس فقط للخبر المصاحب للبيانات، بل أيضاً لتطبيق الأخبار في حد ذاته.

بالنسبة لعملية تنظيف البيانات وتحليلها، استخدمنا إكسل وبرنامج نصي للتنظيف وكذلك برنامج مايكروسوفت إكسل. تمت كتابة تطبيق الأخبار بلغة البرمجة 'Ruby on Rails'. كما استخدم التطبيق جافا سكريبت بشكل مكثف.

وإضافة إلى الخبر العام، تضمنت تغطيتنا تطبيق أخبار تفاعلياً، مما أتاح للقراء فرصة الفهم والعثور على أمثلة داخل تلك المجموعة من البيانات الوطنية الضخمة التي تتعلق بها. ومن خلال تطبيقاتنا الإخبارية، يستطيع القراء إيجاد مدارسهم المحلية - مثلاً المدرسة الثانوية المركزية في نيوآرك بنيوجيرسي (http://goo.gl/HJVCf) - ومن ثم التعرف فوراً على جودة أداء المدرسة في العديد من المناطق. كما يمكنهم النقر على زر يتيح لهم المقارنة بين «المدارس في مناطق أكثر فقراً ومناطق أقل فقراً» (http://goo.gl/WrAli)، والاطلاع الفوري على مدارس ثانوية أخرى ونسبة الفقر بها ومدى استعدادها لتقديم دورات في الرياضيات المتقدمة والمواد التأهيلية لدخول الجامعة (AP) ودورات هامة أخرى. وبالنظر إلى المثال الذي أوردناه، فإن المدرسة الثانوية المركزية تتناقض مع مدرسة ميلبورن الثانوية، فقد أظهر مشروع ''عدم تكافؤ الفرص'' أن 1٪ فقط من طلاب مدرسة ميلبورن يحصلون على وجبات غداء مجانية أو مخفضة السعر، بينما يدرس 72٪ منهم مادة واحدة على الأقل من المواد الجامعية. وفي طرف النقيض الآخر، يحصل 85٪ من طلاب المدرسة الثانوية الدولية على وجبات مجانية أو مخفضة

بينما يدرس 1٪ منهم فقط المواد الجامعية.

ومن خلال هذا المثال، يستطيع القراء استخدام ما يعرفونه من معلومات – مثل مدرسة ثانوية محلية – لفهم ما لا يعرفونه: أي توزيع فرص التعليم ومدى تأثير الفقر على تلك الفرص.

كما قمنا بإضافة هذا التطبيق إلى فيسبوك ليتمكن القراء من الدخول إلى الموقع حيث يقترح تطبيقنا تلقائياً المدارس التي قد تهمهم.

إن معدل زيارات جميع تطبيقاتنا الإخبارية ممتاز ونحن فخورون بالطريقة التي يوضح بها هذا التطبيق أخباراً معقدة، أي أقرب إلى صميم الموضوع، كما يساعد قراءنا على كتابة الأخبار الخاصة بهم.

وكما هو الحال بالنسبة لمعظم المشروعات التي تبدأ باستخدام البيانات الحكومية، فقد تطلبت تلك البيانات الكثير من التنظيف. فعلى سبيل المثال، مع وجود حوالى 30 مادة فقط من المواد الجامعية، أبلغت بعض المدارس أن لديها المئات منها، مما استدعى الكثير من المراجعة والمكالمات الهاتفية مع المدارس لتأكيد تلك البيانات وتصحيحها.

كما أننا عملنا بكل جد كي يتمكن التطبيق من سرد خبرين، أولهما "بعيد" وثانيهما "قريب". بمعنى أنه كان على التطبيق أن يقدم للقارئ صورة وطنية مجردة وواسعة، وهي وسيلة لمقارنة أداء الولايات فيما بينها فيما يتعلق بإتاحة فرص التعليم. إلا أن تقديم تلك الصورة المجردة قد يربك القراء حول مدلول تلك البيانات. كما أننا أردنا أن يتمكن القراء من العثور على مدارسهم المحلية ومقارنتها بالمدارس ذات مستويات الفقر المرتفعة أو المنخفضة في مناطقهم.

وإذا ما تسنى لي تقديم النصح إلى صحفيي البيانات الطامحين والمهتمين بتنفيذ هذا النوع من المشاريع، سأقول إن عليكم الإلمام بالمادة وعليكم أن تكونوا فضوليين! يمكن تطبيق جميع القواعد المطبقة على مختلف أنواع العمل الصحفي هنا. فعليكم الحصول على الحقائق والحرص على إجادة نقل الخبر والتأكد من أن تطبيقكم الإخباري لا يتعارض مع الخبر الذي تكتبونه. وإذا حدث مثل هذا التعارض، فإن أحدهما قد يكون خطأً.

وإذا ما رغبت في تعلم الترميز (كتابة الكود)، فإن أهم خطوة هي البدء. قد

ترغب في التعلم من خلال الدورات التعليمية أو الكتب أو مقاطع الفيديو، ولكن احرص على التوصل إلى فكرة جيدة جداً لمشروع ما مع وجود موعد نهائي لإتمامه. إذا كان في ذهنك خبر لا يمكن نقله إلا عن طريق تطبيق إخباري، فلن يعيقك عدم معرفتك بالبرمجة عن تنفيذ المشروع.

— سكوت كلاين، بروبابليكا

تحقيق لمدة تسعة أشهر حول الصناديق الهيكلية الأوروبية

خلال عام 2010 تعاونت صحيفة فاينانشيال تايمز (http://www.ft.com/intl/eu-funds) مع مكتب الصحافة الاستقصائية (http://bit.ly/bureau-billions) للتحقيق في الصناديق الهيكلية الأوروبية. كان الهدف من التحقيق معرفة المستفيدين من الصناديق الهيكلية الأوروبية والتأكد من عدم إهدار تلك الأموال. إن برنامج الصناديق الهيكلية الذي بلغ حجمه 347 مليار يورو على مدى سبعة أعوام هو ثاني أكبر برنامج للدعم المالي في الاتحاد الأوروبي، وقد ظل البرنامج قائماً عدة عقود؛ ولكن، فيما عدا التوضيحات العريضة والعمومية، كان هناك قليل من الشفافية حول المستفيدين منه. وفي محاولة لإجراء بعض التعديلات في القواعد خلال دورة التمويل الراهنة، تم إجبار السلطات المسؤولة على نشر قائمة بأسماء المستفيدين بما في ذلك مواصفات المشروع وحجم التمويل الذي يتلقاه من الاتحاد الأوروبي والتمويل الوطني.

كان الفريق المختص بالعمل على المشروع يتكون من 12 صحفياً بالإضافة إلى مبرمج متفرغ للعمل تسعة أشهر، بينما استغرق جمع البيانات وحده عدة أشهر.

وقد تمت تغطية المشروع لمدة خمسة أيام في صحيفة فاينانشيال تايمز ومكتب الصحافة الاستقصائية وبرنامج وثائقي في راديو بي. بي. سي، إضافة إلى العديد من الأفلام الوثائقية التلفزيونية.

قبل أن تبدأ العمل على مشروع بهذا الحجم من الجهد، عليك أن تتأكد من صحة نتائج التحقيق ومن أنك ستحصل على أخبار جيدة لم يسبقك إليها أحد.

تم تقسيم المهمة إلى عدد من الخطوات الواضحة.

THE BUREAU OF INVESTIGATIVE JOURNALISM

Follow us 

About the Bureau
Steal our stories
Donate to the Bureau
Subscribe to our Newsletter
Contacting the Bureau

Home All Stories Corporate Watch Health Human Rights Open Society

Europe's Hidden Billions



نموذج 3–2. التحقيق في الصناديق الهيكلية الأوروبية
(صحيفة فاينانشال تايمز ومكتب الصحافة الاستقصائية)

1– تحديد من الذي يحتفظ بالبيانات وكيفية الاحتفاظ بها

لدى الإدارة العامة للمفوضية الأوروبية بوابة (http://bit.ly/ec–portal) مفتوحة على مواقع السلطات الإقليمية التي تنشر البيانات. كنا نعتقد أن المفوضية لديها قاعدة بيانات شاملة لبيانات المشروع يمكننا إما الوصول إليها مباشرة، أو الحصول عليها عن طريق تقديم طلب حرية المعلومات. ولكننا لم نجد قاعدة بيانات مفصلة بالمستوى الذي نحتاج إليه. وأدركنا على الفور أن عدداً كبيراً من الروابط التي وفرتها المفوضية معطوبة وأن معظم السلطات قد نشرت البيانات في صيغة ملفات PDF بدلاً من الصيغ القابلة للتحليل مثل CSV أو XML.

عمل فريق مكون من 12 شخصاً للتعرف على أحدث البيانات وجمع الروابط في جدول بياني واحد استخدمناه للتعاون فيما بيننا. وبما أن خانات البيانات

لم تكن موحدة، (مثلاً كُتبت العناوين بلغات مختلفة كما أن بعض مجموعات البيانات استخدمت عملات مختلفة وشمل بعضها تحليلاً لتمويل الاتحاد الأوروبي والتمويل الوطني)، كان علينا أن ندقق قدر المستطاع في ترجمة خانات البيانات المتاحة في كل مجموعة ووصفها.

2- تحميل البيانات وإعدادها

شملت الخطوة التالية تحميل كافة الجداول البيانية وملفات PDF وفي بعض الأحيان استخراج البيانات الأصلية من الويب.

كان لا بد إذن من توحيد مقاييس كل مجموعة بيانات. كانت أكبر مهامنا تتمثل في استخراج البيانات من ملفات PDF التي بلغ طول بعضها مئات الصفحات. وقد أنجزنا معظمها باستخدام برنامجي UnPDF وABBY FineReader اللذين يسمحان باستخراج البيانات في صيغ CSV أو إكسل.

كما تطلب العمل التفقد والمراجعة للتأكد من أن أدوات استخراج البيانات من ملفات PDF قد نقلت البيانات بالطريقة الصحيحة. وقد تم ذلك عن طريق التنقية والتصنيف وجمع الإجمالي (وهذا للتأكد من مطابقتها للمادة المطبوعة في ملفات PDF).

3- إنشاء قاعدة بيانات

تولى مبرمج الفريق إنشاء قاعدة بيانات SQL حيث تم استخدام كل من الملفات المعدة كمكونات لبناء قاعدة بيانات SQL شاملة. ومن خلال عملية يومية يتم تحميل كافة ملفات البيانات الفردية في قاعدة بيانات SQL كبيرة يمكن استخدامها للبحث عن طريق واجهتها أو بإدخال كلمات بحثية.

4- إعادة التدقيق والتحليل

قام الفريق بتحليل البيانات بطريقتين أساسيتين:

عن طريق واجهة قاعدة البيانات

استلزمت هذه الطريقة كتابة كلمات بحثية معينة (مثال: "تبغ"، "فندق"، "شركة

أ") في محرك البحث. وبمساعدة برنامج ترجمة جوجل، والذي تم تضمينه في خاصية البحث بقاعدة البيانات، تتم ترجمة تلك الكلمات البحثية إلى 21 لغة مما مكننا من الحصول على النتائج المطلوبة التي يمكن تحميلها ليتمكن الصحافيون من إجراء المزيد من البحث حول المشروعات التي تهمهم.

عن طريق التحليل الكلى باستخدام قاعدة البيانات بأكملها

من حين لآخر نقوم بتحميل مجموعة بيانات كاملة يمكننا بعد ذلك أن نحللها (مثلاً باستخدام كلمات بحثية أو جمع بيانات المحصلة حسب البلد أو الإقليم أو نوع الإنفاق أو عدد المشروعات حسب المستفيدين منها.. إلخ).

لقد حددت هاتان الطريقتان معالم الخبر وكذلك من خلال البحث الميداني والمكتبي.

استغرقت عملية إعادة التحقق من صحة البيانات وقتاً طويلاً (من خلال جمع المحصلات ومراجعة ما تم تخصيصه حسبما ما تقوله السلطات). وتبين أن إحدى المشاكل الرئيسية كانت أن السلطات تقوم في معظم الأحيان بالكشف فقط عن كم التمويل المقدم من "الاتحاد الأوروبي والتمويل الوطني". وبموجب أحكام الاتحاد الأوروبي يسمح لكل برنامج بتمويل نسبة مئوية معينة من التكلفة الكلية باستخدام تمويل الاتحاد الأوروبي الذي يتم تحديده، على مستوى البرامج، وفقاً لما يُطلق عليه "معدل التمويل المشترك". ويتكون كل برنامج (مثل برنامج المنافسة الإقليمية) من عدة مشروعات. وعلى مستوى المشروعات، يمكن لأحدها أن يحصل على 100٪ من تمويل الاتحاد الأوروبي، بينما قد لا يتلقى مشروع آخر أي تمويل. وما دامت المشروعات في مجموعة واحدة، فلن يزيد تمويل الاتحاد الأوروبي على مستوى البرامج عن معدل التمويل المشترك المتفق عليه.

وقد عنى ذلك أنه كان علينا التحقق من كل مبلغ من تمويل الاتحاد الأوروبي الذي ذكرناه في خبرنا مع ذكر الشركات المستفيدة المعنية.

— سينثيا أومورشو، صحيفة فاينانشيال تايمز

انهيار منطقة اليورو

نحن نقوم بتغطية موضوع انهيار منطقة اليورو (http: //on.wsj.com /tY-M82O) بكل ما يتعلق به من تفاصيل. ويتمثل الموقف الدرامي في المواجهات بين الحكومات وفقدان المدخرات الشخصية وردود الأفعال من قادة دول العالم وإجراءات التقشف ثم الاحتجاجات المقامة ضدها. تنشر صحيفة وول ستريت جورنال يومياً رسوما بيانية عن الوظائف المفقودة وانخفاض إجمالي الناتج المحلي وركود الأسواق العالمية. كل هذا يتزايد بشكل يثير الذهول.

وبينما نخرج من اجتماع دعا إليه محررو الصفحات الأولى لمناقشة الأفكار حول تغطيات نهاية العام، وجدت نفسي أتساءل كيف لنا أن نتعايش مع هذه الأوضاع.

هل الأمر مماثل لما حدث عام 2008 عندما تم تسريحي من العمل وتوالت الأخبار السيئة؟ تحدثنا عن الوظائف والعمل والمال كل ليلة أثناء تناول العشاء، وكدت أنسى أن ذلك قد يثير استياء ابنتي. حاولت إنكار ذلك الخوف الذي سبب لي آلام الرقبة والقلق الذي صار يشكل ضغطاً هائلاً على صدري. هل هذا يشبه الوضع الراهن للأسر في اليونان وإسبانيا؟

عدت وماشيت مايك ألين – محرر الصفحة الأولى – وهو في طريقه إلى مكتبه وعرضت عليه فكرة سرد الأزمة من خلال الأسر في منطقة اليورو، بدءاً بالنظر إلى البيانات والبحث في الخلفيات الديمغرافية لتساعدنا في فهم التركيبة الأسرية، ثم بالاستعانة بالصور والمقابلات والتسجيلات لأجيال تلك الأسر. وقد استخدمنا صوراً وأصواتاً معبرة علاوة على البيانات المتوفرة.

عندما عدت إلى مكتبي كتبت نبذة مختصرة وابتكرت شعاراً.

في الأسابيع الثلاثة التالية قمت بمتابعة الأرقام بتقصي إحصاءات الزواج والوفيات وحجم الأسر والإنفاق على الرعاية الصحية. كما نظرت في الترتيبات المعيشية ومعدلات الطلاق واطلعت على دراسات حول معدلات الرفاهية والادخار وتصفحت كذلك أقسام الإحصاءات القومية واتصلت بمكتب الأمم المتحدة المعني بالسكان وصندوق النقد الدولي والإحصاءات الأوروبية (Eurostat)

In 1993 the Maastricht treaty bound 17 countries with distinctly different cultures and centuries of history into one entity; the European Union.

Fifteen years later, in 2009, on the heels of a global financial meltdown, fears of a debt crisis began to spread from Iceland to Greece to Germany.

In 2011, the EU faces financial and political turmoil on an unprecedented scale, austerity measures, looming bailouts and financial uncertainty.

What is it like to be living in the midst of instability in a country where your family has roots that go back generation upon generation, paying bills with a currency that has been changing hands for less than a decade, entangled in a group of economies on the verge of default?

The Wall Street Journal spoke to families in six of those countries to find out:

Families of countries

Interactive portraits, voices and profiles of families from the euro zone crisis



نموذج 3–3. انهيار منطقة اليورو: نبذة مختصرة (صحيفة وول ستريت جورنال)

والمنظمة الأوروبية للتعاون الاقتصادي والتنمية ((OECD حتى عثرت أخيراً على خبير اقتصادي قضى حياته العملية في رصد الحياة الأسرية وقام بتعريفي بإحدى الباحثات في مجال تكوين الأسرة فأرشدتني إلى وثائق حكومية بشأن الموضوع الذي أعمل عليه.

وبمساعدة المحرر سام إنريكيز قمنا بحصر البلدان وجمعنا فريقاً لمناقشة النهج المرئي وأي الصحافيين يمكنه نقل الكلمات والتسجيل الصوتي والخبر. بدأ مات كريغ – محرر الصور للصفحة الأولى – العمل بالبحث عن المصورين. كما قام مات موراي – نائب مدير التحرير للتغطيات العالمية – بإرسال مذكرة لرؤساء المكاتب طالباً المساعدة من الصحافيين. (كان ذلك مهماً: الحصول على موافقة الرؤساء).

إدراكا بأن الأولوية للبيانات، توليت تصدير البيانات كل صباح إلى الجداول البيانية وأعددت رسوماً بيانية لمعرفة التوجهات: تقلص المدخرات، اختفاء المعاشات، عودة الأمهات إلى العمل، تكاليف الرعاية الصحية، بالإضافة إلى ديون الحكومة والبطالة. أما في فترات ما بعد الظهر، فقد كنت أنظر في مجموعات

البيانات ثم أقارن البلدان بعضها ببعض للعثور على موضوعات للأخبار.

داومت على هذا العمل لمدة أسبوع قبل أن أفقد طريقي وأبدأ التشكيك في نفسي. ربما كان هذا نهجاً خاطئاً، أو ربما لم يكن الأمر متعلقاً بالبلدان، ولكنه يتعلق بالآباء والأمهات والأطفال والأجداد.

كانت البيانات تنمو ثم تتقلص. كنت أعمل أحياناً لساعات طويلة على جمع المعلومات لأدرك في نهاية المطاف إنها لا تفيدني في شيء أو إنني بحثت في مجموعة أرقام خطأً كلياً. أحياناً كانت البيانات أقدم من اللازم.

Dec 15, 2011 5:37 PM ITALY_onlychildren.xlsx 123 KB XML File

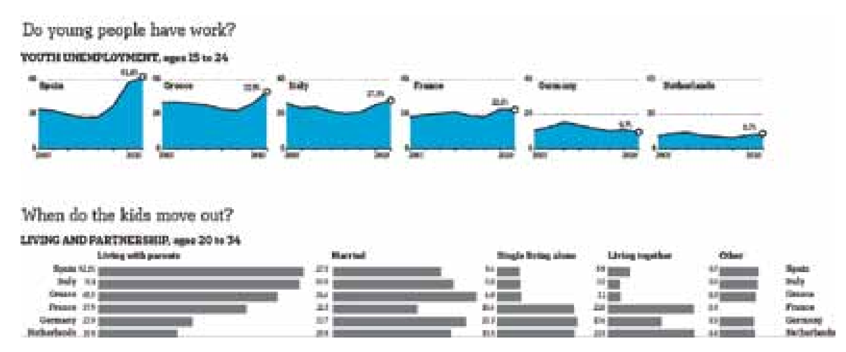
Distribution of	households by	number of children, 2007 (2)		
	1 child	2	3	4+
Sweden	43.3	40.6	12.8	3.[illegible]
Finland	42.7	39.2	13.5	4.[illegible]
Denmark	41.3	43.4	12.5	2.[illegible]
Netherlands	38.8	42.7	14.1	4.[illegible]
France	45.3	39.9	11.7	3.[illegible]
Germany	48.6	39.5	9	[illegible]
Austria	50.1	37.2	10.2	2.[illegible]
Belgium	44.5	36.8	13.7	[illegible]
Luxembourg	44.8	46	8.1	1.[illegible]
Ireland	43.8	35.2	16	[illegible]
Italy	55.2	37.9	6.1	0.[illegible]
Spain	55.2	39.9	3.9	0.[illegible]
Portugal	61.4	33.7	4	[illegible]
Greece	46.4	47.9	4.3	1.[illegible]
Cyprus	42.5	46.8	8.5	2.[illegible]
Hungary	49.5	36.9	10.5	3.[illegible]
Estonia	58	32.9	7.5	1.[illegible]
Latvia	62.8	29.5	5.8	1.[illegible]
Lithuania	59.7	31.4	6.8	2.[illegible]
Slovenia	49.7	41.5	7.2	1.[illegible]
Slovakia	53.7	36	8.3	[illegible]
Poland	53.5	35.2	8.6	2.[illegible]
EU-25	49.5	38.9	9	2.[illegible]
EU-15	48.7	39.5	9.2	2.[illegible]
NMS	[illegible]	[illegible]	[illegible]	[illegible]

نموذج 3–4. قد يستغرق اختبار جدوى مجموعة البيانات وقتاً طويلاً (سارة سلوبان)

ومع ازدياد حجم البيانات مرة أخرى أدركت أن هناك أسئلة ظلت بلا إجابات وأنني لم أتمكن من فهم الأُسر.

كان لزاماً عليّ أن أصل إلى نتيجة، لذا قمت بإعداد سلسلة سريعة من الرسوم البيانية عبر برنامج Illustrator وبدأت بتنظيمها وتعديلها.

وبعد إنجاز الرسوم البيانية اتضحت لنا صورة متماسكة للأسر.

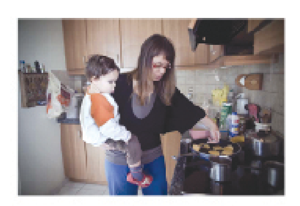


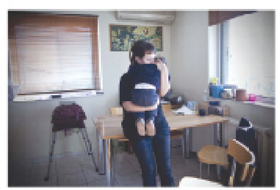
نموذج 3–5. الرسوم البيانية المرئية: محاولة فهم منطقي للتوجهات والأنماط المخبأة ضمن مجموعات البيانات (سارة سلوبان)

بدأنا في العمل بعد استدعاء الصحافيين وتسليمهم الرسوم البيانية والعرض العام وتوجيه دعوة مفتوحة لهم للبحث عن الأخبار ذات المغزى والتي يمكنها أن تقرب صورة الأزمة لقرائنا. احتجنا إلى أسرة صغيرة في أمستردام وأسر أخرى كبيرة في إسبانيا وإيطاليا. كنا نود الاستماع إلى العديد من الأجيال لمعرفة تأثير التاريخ الشخصي على تشكيل الردود.

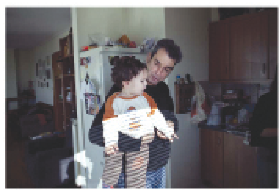
ومن ذاك الوقت فصاعداً كنت أستيقظ مبكراً لمراجعة بريدي الإلكتروني لمراعاة فروق التوقيت. وقد عاد الصحافيون وفي جعبتهم مواضيع جيدة وملخصات ومفاجآت لم أكن أتوقعها.

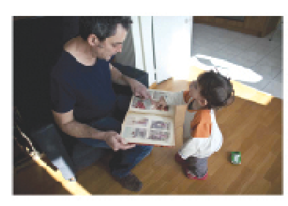
فيما يتعلق بالتصوير كنا ندرك أننا نود التقاط صور للأجيال. كان "مات" يرى أن على مصوريه متابعة أفراد الأسر يوماً كاملاً لتصوير حياتهم اليومية. ومن ثم فقد اختار صحافيي المرئيات الذين غطوا مهمات عالمية تشمل الأخبار وحتى ميادين الحروب. أراد "مات" أن تنتهى كل مهمة تصوير على مائدة العشاء. أما سام فقد اقترح إضافة قوائم الطعام.

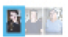




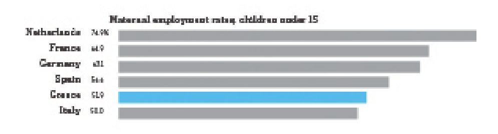




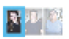




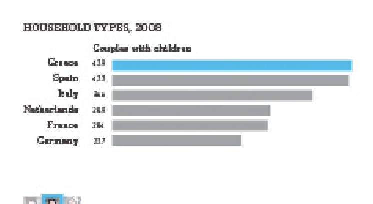
"In no way am I happy with my quality of life because the moment we brought a child into the world my husband became unemployed. The result was that I couldn't enjoy motherhood as I immediately had to return to work.



Katerina's husband, Konstantinos, 50, is unemployed afterlosing both his jobs: at a clothes warehouse and as a night watchman at the Ancient Agora archaeological site. He's no longer entitled to unemployment benefit.



....A woman is exonerated from not working, it is something ordinary. But for a man it is something harsh and this is yet an additional strain at home.



نموذج 3–6. الأرقام تعبر عن الناس: تكمن قيمة البيانات في الأخبار الفردية التي تمثلها (صحيفة وول ستريت جورنال)

ومن هنا كان علينا الانتظار لمعرفة ما الذي تسرده الصور للاطلاع على ما قالته تلك الأسر. قمنا بتصميم شكل العمل التفاعلي واقتباس لوحات لونية من إحدى روايات "تان تان". وبعد استكمال العملية وحصولنا على القصة المصورة، أضفنا بعض الرسوم البيانية الأصلية، أضفناها فقط بما يكفي لضبط كل قصة وتقوية مضمونها. أما البيانات فقد شكلت وقفات فاصلة في القصة أضفت نوعا من التغيير.



نموذج 3–7. الحياة في منطقة اليورو (صحيفة وول ستريت جورنال)

في النهاية فإن البيانات هي الأشخاص والصور والقصص، وكانت بمثابة إطار لكل رواية أو خطاب والقوة الدافعة للتوتر بين الدول.

وعندما حان وقت النشر، قبيل بداية العام الجديد بينما كنا جميعاً نتأمل ما هو آتٍ، كنت قد تعرفت على أسماء كافة أفراد الأسر، ولا أزال أتساءل كيف حالهم الآن. وإذا لم يكن هذا يشبه مشاريع البيانات، فلن أبالي لأننا استطعنا مشاركة قرائنا في تلك اللحظات التي قمنا بتوثيقها في مشروع "الحياة في منطقة اليورو" حيث يجري الحديث مع أسرتك أثناء تناول الطعام عن العمل والحياة. وقد تحقق كل هذا بفضل فهم البيانات.

—— سارة سلوبان، صحيفة وول ستريت جورنال.

تغطية الخزانة العامة على الموقع الإلكتروني OpenSpending.org

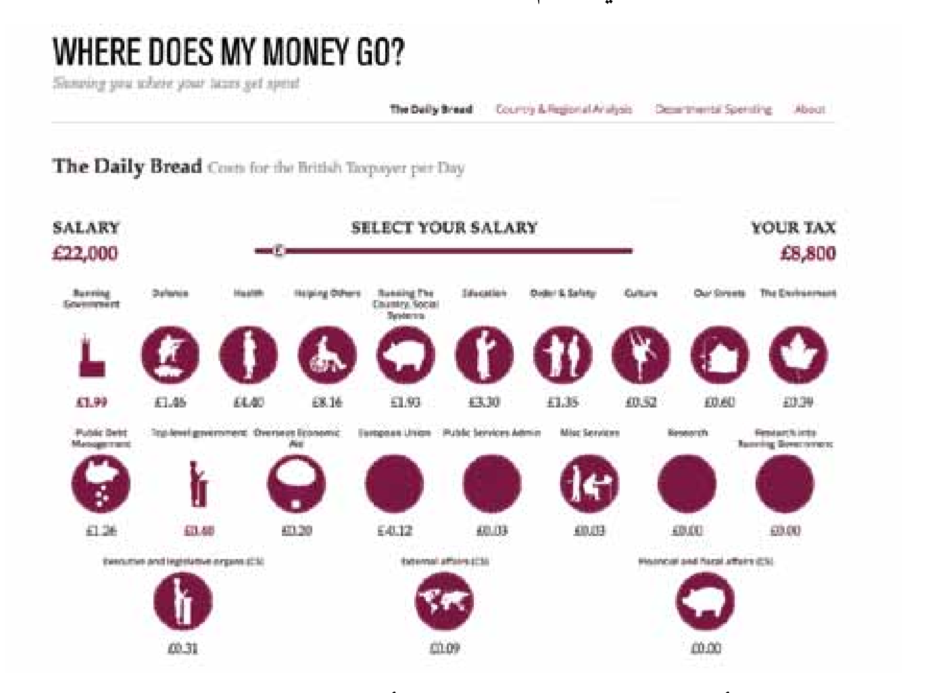
في عام 2007 ذهب جوناثان إلى «مؤسسة المعرفة المفتوحة» مقترحاً مشروعاً من صفحة واحدة يدعى 'أين تذهب نقودي؟' (http://www.wheredoesmy-moneygo.org/). كان المشروع يهدف إلى إطلاع المواطنين البريطانيين على كيفية إنفاق المال العام، كما كان مقدمة إثبات مفاهيم لمشروع أكبر يمثل الإعلام العام مرئياً بناء على العمل الرائد الذي قام به معهد آيسوتايب Isotype التابع لـ 'أوتّو وماري نيوراث Otto and Marie Neurath's' خلال الأربعينيات من القرن الماضي.



نموذج 3–8. أين تذهب نقودي؟ (مؤسسة المعرفة المفتوحة)

لقد أتاح مشروع «أين تذهب نقودي" للمستخدمين فرصة الاطلاع على البيانات العامة من خلال مجموعة واسعة من المصادر باستخدام أدوات مفتوحة المصدر وسهلة الاستعمال. وقد نلنا عرضاً للمساعدة في تطوير نموذج أولي للمشروع وتسلمنا لاحقاً تمويلاً من إذاعة 4IP التابعة للقناة الرابعة لتحويل المشروع إلى تطبيق ويب كامل. وقام خبير تصميم المعلومات دافيد ماكاندليس (من خلال كتابه 'روعة المعلومات': http://www.informationisbeautiful.net) بطرح العديد من تصاميم البيانات التي ساعدت الناس على التعامل مع الأرقام الضخمة،

بما في ذلك «التحليل الوطني والإقليمي“ الذي يوضح كيفية صرف الأموال العامة على مختلف أجزاء البلاد، و»المعيشة اليومية» (http://wheredoesmymon-eygo.org/dailybread.html) الذي يوضح للمواطنين تفاصيل صرف الأموال المجموعة من الضرائب في اليوم الواحد بالجنيه والبنس.



نموذج 3–9. أداة حساب الضرائب في مشروعي "أين تذهب نقودي" و"المعيشة اليومية" (مؤسسة المعرفة المفتوحة)

في ذاك الوقت، كان المشروع يرتكز على قاعدة بيانات «نظام المعلومات الموحدة على الإنترنت» (أو COINS) (http://data.gov.uk/dataset/coins) وهي قاعدة البيانات الأكثر شمولاً وتفصيلاً للمصروفات الحكومية بالمملكة المتحدة. أثناء عملنا مع ليزا إيفانز (قبل انضمامها إلى فريق مدونة بيانات الغارديان)، وجوليان تود وفرانسيس إرفينج (الذي اشتهر بسبب موقع ScraperWiki) ومارتن روزنبام (بي. بي. سي.) وآخرين، قدمنا العديد من الطلبات للحصول على البيانات، ولكن تم رفض أغلب تلك الطلبات. (قامت ليزا بتوثيق التقرير المفصل جزئيا في العمود الهامشي ”استخدام قانون حرية المعلومات لفهم

الإنفاق" في صفحة 120).

وعندما تم نشر البيانات أخيراً في منتصف عام 2010، اعتبرها المدافعون عن الشفافية ثورة في هذا المجال. وقد حصلنا على امتياز مسبق للوصول إلى البيانات وتحميلها إلى تطبيقنا على الويب، كما حظينا باهتمام إعلامي بالغ عند الإعلان عن ذلك. وفي يوم إطلاق الخدمة ظهر العديد من الصحافيين على قناة IRC الخاصة بنا لمناقشة جوانب الخدمة والاستفسار عنها وعن وكيفية فتح الملفات واستكشافها (بلغ حجم الملفات العشرات من الغيغابايتات). وبينما ادعى بعض النقاد أن تلك الخدمة الضخمة كانت في غاية التعقيد حتى أنها كانت عبارة عن "غموض من خلال الشفافية» (http://bit.ly/archive-silicon)، دقق الكثير من الصحافيين الشجعان في البيانات ليقدموا لقرائهم صورة غير مسبوقة لكيفية إنفاق الأموال العامة. وقد قامت صحيفة الغارديان بالتدوين المباشر على مدونتها (http://bit.ly/guardian-coins) عن إطلاق الخدمة، بالإضافة إلى تغطية العديد من وسائل الإعلام الأخرى التي قدمت تحليلاتها لنتائج البيانات.

لم يمض الكثير من الوقت حتى بدأنا في تلقي الطلبات والاستفسارات حول تنفيذ مشروعات مشابهة في بلدان أخرى في العالم. وبعد وقت قصير من إطلاق مشروع OffenerHaushalt (http://offenerhaushalt.de) - وهو نسخة أخرى من المشروع خاصة بميزانية الدولة الألمانية ابتكرها فريدريك ليندنبيرج - قمنا بإطلاق مشروع OpenSpending (http://openspending.org/) - النسخة الدولية من المشروع - الذي يهدف إلى مساعدة مستخدميه على رصد الإنفاق العام حول العالم مثلما ساعدهم مشروع Open Street Map على رسم خريطة للخصائص الجغرافية. وقد قمنا بتطبيق تصاميم جديدة - بمساعدة الموهوب غريغور آيش - بنيت جزئياً على التصاميم الأصلية التي أعدها دافيد ماكاندليس.

أما بالنسبة لمشروع OpenSpending، فقد عملنا على نطاق واسع مع صحافيين للحصول على بيانات الإنفاق وتمثيلها وتفسيرها وتقديمها للجمهور. هذا المشروع في المقام الأول عبارة عن قاعدة بيانات ضخمة وقابلة للبحث للإنفاق العام على المستويات العليا للمعلومات المتعلقة بالميزانية وعلى مستوى الإنفاق الفعلي للأموال العامة. وعلى هذا الأساس تم بناء سلسلة من التوضيحات



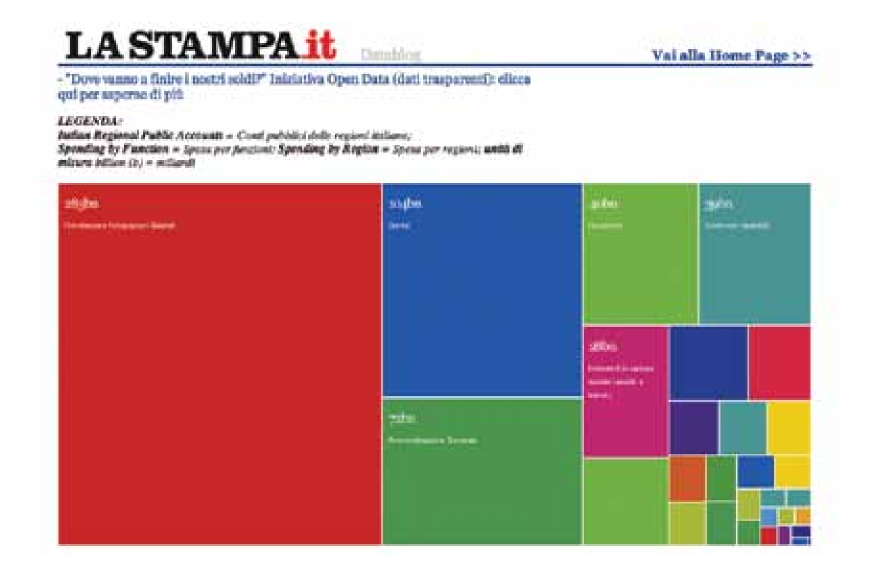
نموذج 3–10. OffenerHaushalt "النسخة الألمانية من مشروع 'أين تذهب نقودي؟"
(مؤسسة المعرفة المفتوحة)

المرئية غير المسبوقة مثل Treemap وBubbletree. يستطيع كل فرد تحميل البرنامج في بيانات المجلس المحلي ومن ثم إنتاج التوضيحات المرئية من خلاله.

وبينما كنا نظن في البدء أن الطلب على بعض من توضيحاتنا المرئية المتطوّرة سيزداد، أدركنا بعد التحدث إلى المؤسسات الصحفية أن هناك المزيد من الاحتياجات الأساسية التي لا بد من تلبيتها أولاً، مثل القدرة على تثبيت جداول البيانات المتحركة في التدوينات. ولأننا حريصون على تشجيع المؤسسات الصحفية على تلبية حاجة الجمهور للوصول إلى البيانات إلى جانب الأخبار، قمنا ببناء أداة لهذا الغرض أيضاً.

تزامن أول نشر كبير لنا مع مهرجان الصحافة الدولي الأول في بيروجيا، حيث تعاونت مجموعة من المطورين والصحفيين وموظفي الحكومة لتحميل بيانات الحكومة الإيطالية على منصة OpenSpending، مما قدم رؤية غنية عن تقسيم الإنفاق الحكومي على الإدارات المركزية والإقليمية والمحلية. وقد تمت

تغطية هذا المشروع من قبل صحف ‘إل فاتو كوتيديانو’ (-http: / /bit.ly /ilfat to-spending) و‘إل بوست’ (http: / /bit.ly /ilpost-spending) و‘لا ستامبا’ (http: / /bit.ly /lastampa-spending) و‘ريبابليكا’ (-http: / /bit.ly /repub blica-spending) و‘وايرد إيطاليا’ (http: / /bit.ly /wired-italy-spending) والغارديان (http: / /bit.ly /guardian-italy-spending).



نموذج 3–11. النسخة الإيطالية من مشروع ’أين تذهب نقودي؟‘
(صحيفة لا ستامبا)

في عام 2011، عملنا مع مشروع ’انشر ما تموّل‘ (-http: / /www.pub lishwhatyoufund.org/) ومعهد التنمية لما وراء البحار (.http: / /www.odi org.uk/) لرصد تمويل المساعدات لأوغندا خلال الفترة من 2003 إلى 2006. كان هذا أمراً جديداً لأنك للمرة الأولى تتمكن من رؤية تدفق تمويل المساعدات إلى جانب الميزانية العامة، مما يتيح لك معرفة مدى تطابق أولويات المانحين مع أولويات الحكومات. وقد ظهرت بعض النتائج المثيرة للاهتمام، منها على سبيل المثال أن برامج مكافحة مرض الإيدز وبرامج تنظيم الأسرة يموّلان كلياً تقريباً بواسطة المانحين الأجانب، وقد قامت صحيفة الغارديان بتغطية الموضوع (http: / /bit.ly /guardian-uganda-viz).

كما عملنا كذلك مع المنظمات غير الحكومية وجماعات الدفاع عن الحقوق لمقارنة بيانات الإنفاق مع مصادر المعلومات الأخرى. وعلى سبيل المثال، زودتنا منظمة الخصوصية الدولية بقائمة ضخمة لشركات المراقبة التكنولوجية وقائمة بالوكالات المشاركة في المعرض الدولي التجاري الشهير، والذي يُعرف باسم 'حفل المتنصتين' باللغة العامية. ومن خلال المقارنة المنظمة لأسماء الشركات مع مجموعات بيانات الإنفاق، تمكنّا من تحديد أي الشركات لديها عقود مع الحكومة، وهذا ما يمكننا تتبعه لاحقا بطلبات حرية المعلومات. وقد قامت صحيفة الغارديان بتغطية هذا الخبر (http://bit.ly/guardian-surveillance).

نسعى حالياً لزيادة الوعى بالسياسات المالية لدى الصحافيين والجمهور كجزء من مشروع يُدعى 'أخبار الإنفاق' (http://bit.ly/ss-faq) والذي يسمح للمستخدمين بربط بيانات الإنفاق العام بأخبار تتعلق بالإنفاق العام للاطلاع على الأرقام المصاحبة للأخبار والأخبار الكامنة وراء الأرقام.

ومن خلال عملنا في هذا المجال، تعلمنا ما يلي:

- عادة ما تجد أن الصحافيين غير معتادين على التعامل مع البيانات الخام، كما أن الكثيرين منهم لا يعتبرها أساساً هاماً لإعداد تقاريرهم، فاستخلاص الأخبار من المعلومات الخام لا يزال فكرة جديدة نسبيا.
- تستغرق عملية تحليل البيانات وفهمها الكثير من الوقت حتى مع توفر المهارات اللازمة. ومن الصعب دمج هذه العملية مع دورة إخبارية قصيرة الأجل لذا غالباً ما تُستخدم صحافة البيانات في المشاريع الاستقصائية طويلة الأجل.
- غالباً ما تكون البيانات التي تصدرها الحكومات ناقصة أو قديمة. وفي أغلب الأحيان، لا يمكن استخدام قواعد البيانات الحكومية في أغراض التحقيق دون إضافة المزيد من المعلومات المحددة التي يتم طلبها عن طريق حق الحصول على المعلومات (FOI).
- في الغالب تملك مجموعات الدفاع عن الحقوق والعلماء والباحثون المزيد من الوقت والموارد لإجراء البحوث المكثفة والمدعومة بالبيانات، أكثر مما هو متاح للصحفيين. ومن المفيد جداً الانضمام إليهم والعمل في شكل فرق.

— لوسي تشامبرز وجوناثان غراي، مؤسسة المعرفة المفتوحة

الانتخابات البرلمانية الفنلندية وتمويل الحملات الانتخابية

جرت في الأشهر الأخيرة العديد من المحاكمات المتعلقة بتمويل الحملات الانتخابية خلال الانتخابات العامة بفنلندا لعام 2007.

عقب انتخابات عام 2007، اكتشف الإعلام أن الساسة قد تجاهلوا القوانين الملزمة بنشر تفاصيل تمويل الحملات الانتخابية. وقد استخدم تمويل الحملات الانتخابية بشكل أساسي للمحسوبية وشراء الامتيازات من السياسيين الذين امتنعوا حينها عن الكشف عن مصادر تمويل حملاتهم بموجب القانون الفنلندي.

صار القانون أكثر تشدداً بعد الكشف عن تلك الممارسات. فبعد الانتخابات العامة التي جرت في مارس عام 2011، قرر هيلسنغن سانومات الحصول على كافة البيانات المتوفرة حول تمويل الحملات الانتخابية. يشترط القانون الجديد الإعلان عن أي تمويل للحملات الانتخابية فيما عدا التبرعات التي تقل عن 1500 يورو.

1- البحث عن البيانات والمطورين

منذ مارس عام 2011، تولى هيلسنغن سانومات تنظيم ملتقى HS Open للصحافيين والمبرمجين، إذ تمت دعوة مبرمجين فنلنديين وصحافيين ومصممي غرافيكس إلى قبو المبنى. وتم تقسيم المشاركين إلى مجموعات من ثلاثة أشخاص لكل مجموعة حيث تم تشجيعهم على تطوير التطبيقات والتوضيحات المرئية. بلغ عدد المشاركين حوالي 60 مشاركاً في كلٍ من الفعاليات الثلاث التي أقيمت حتى الآن. وقد قررنا أن يكون تمويل الحملات الانتخابية هو مركز اهتمامنا في الملتقى الثاني في مايو عام 2011.

من المعروف أن السلطة المكلفة بالاحتفاظ بسجلات تمويل الحملات الانتخابية هي مكتب التدقيق المحاسبي الوطني بفنلندا. كان هذا الجزء السهل، فقد قام رئيس المكتب الإعلامي جاكو هامونين بإنشاء موقع إلكتروني يتيح الوصول إلى قاعدة البيانات الخاصة بتمويل الحملات الانتخابية. وقد قام مكتب التدقيق بتنفيذ مطلبنا بعد شهرين فقط من تقديم الطلب.

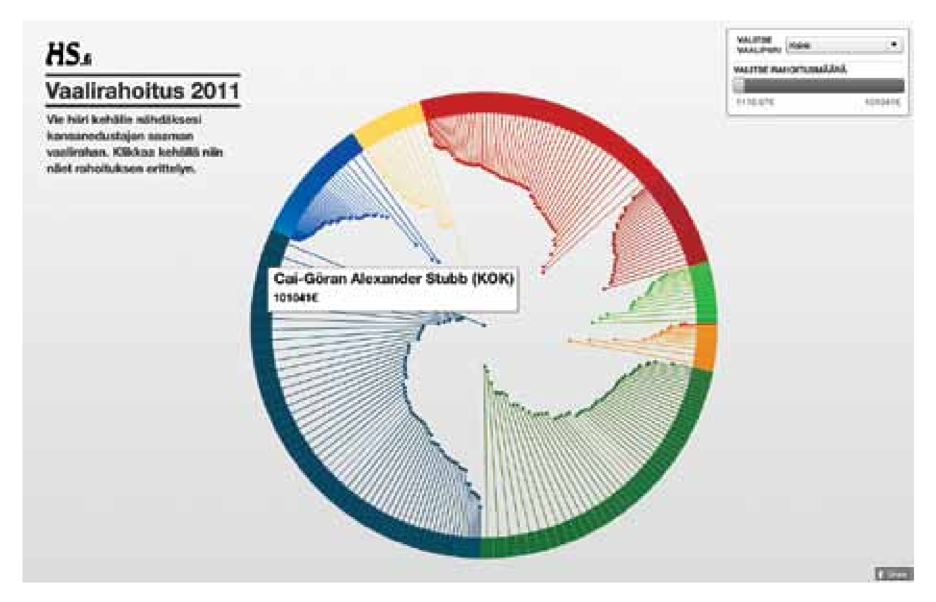
يتولى موقع vaalirahoitus.fi (http://www.vaalirahoitus.fi/) تزويد

الناس والصحافة بالمعلومات عن تمويل الحملات لكل عملية انتخابية من الآن فصاعداً.

2- طرح أفكار جديدة

توصّل المشاركون في ملتقى HS Open الثاني إلى عشرين نموذجا أوليا مختلفا عما يجب فعله بالبيانات. ويمكنكم الاطلاع على كافة النماذج الأولية على موقعنا (http://bit.ly/hs-prototype) (النص باللغة الفنلندية).

قالت جان بيلتولا - وهي باحثة في مجال المعلومات الحيوية - إن بيانات تمويل الحملات الانتخابية تبدو كبيانات الجينات التي يجرون أبحاثهم حولها، كونها تحتوي على العديد من العناصر المترابطة. توجد في مجال المعلومات الحيوية أداة مفتوحة المصدر تسمى Cytoscape (http://www.cytoscape.org/) تُستخدم لرصد تلك العناصر المترابطة. لذا حللنا البيانات باستخدام Cytoscape وحصلنا على نموذج أولي مثير للاهتمام.



نموذج 3–12. تمويل الانتخابات (هيلسنغن سانومات)

3- تنفيذ الفكرة على الورق وعلى الإنترنت

ينص قانون تمويل الحملات الانتخابية على أن نواب البرلمان المنتخبين يجب أن يعلنوا عن التمويل الخاص بحملاتهم بعد شهرين من إجراء الانتخابات. وهذا يعني عملياً أننا نحصل على البيانات الحقيقية في منتصف شهر يونيو. وفي خلال الملتقى، حصلنا فقط على البيانات من النواب الذين تمكنوا من إعدادها قبل نهاية الفترة المحددة.

كانت هناك أيضاً مشكلة تتعلق بصيغة البيانات، إذ قام المكتب الوطني للمراجعة بإعداد البيانات في اثنين من ملفات ((CSV، أحدهما شمل إجمالي ميزانية الحملات الانتخابية، والآخر شمل قائمة بأسماء جميع المانحين. كان علينا إدماج الملفين ثم إنشاء ملف مكون من ثلاثة أعمدة: الجهة المانحة، المستلم والقيمة المدفوعة.

في حال قيام السياسي بتمويل نفسه ذاتياً، نظهر ذلك في بياناتنا كما يلي: السياسي (أ) تبرع بقيمة "س يورو" للسياسي (أ). ورغم أن هذا يبدو أمراً بديهياً، ألا إنه ضروري لتشغيل نظام سايتوسكيب.

وبعد اكتمال عملية تنقيح وإعادة تشكيل البيانات، قمنا بإدخالها في نظام سايتوسكيب، ثم تولت إدارة الرسوم البيانية عملية تكبيرها في شكل رسوم بيانية موسعة.

في نهاية المطاف، قمنا بإعداد توضيح مرئي رائع على موقعنا: (http://www.vaaliraha.com/). وهذا لم يكن رسماً بيانياً للتحليل الشبكي، إذ إننا أردنا أن نوفر للجمهور وسائل سهلة لاطلاعه على حجم التمويل للدعاية الانتخابية وما هي الجهات التي توفره. يشير القسم الأول إلى نصيب كل نائب برلماني من التمويل الدعائي. وعندما تقوم بالنقر على اسم النائب، تحصل على تفاصيل التمويل المقدم له أو لها. كما يمكنك إبداء رأيك من خلال التصويت لصالح أو ضد أي من المانحين. وقد قام كلٌ من جوها روفينين وجوكا كوكو من وكالة إعلانات تسمى ساتوما بإعداد هذا التوضيح المرئي.

تستخدم النسخة الإلكترونية من التوضيح المرئي لتمويل الدعاية الانتخابية نفس البيانات المستخدمة في التحليل الشبكي.

4- نشر البيانات

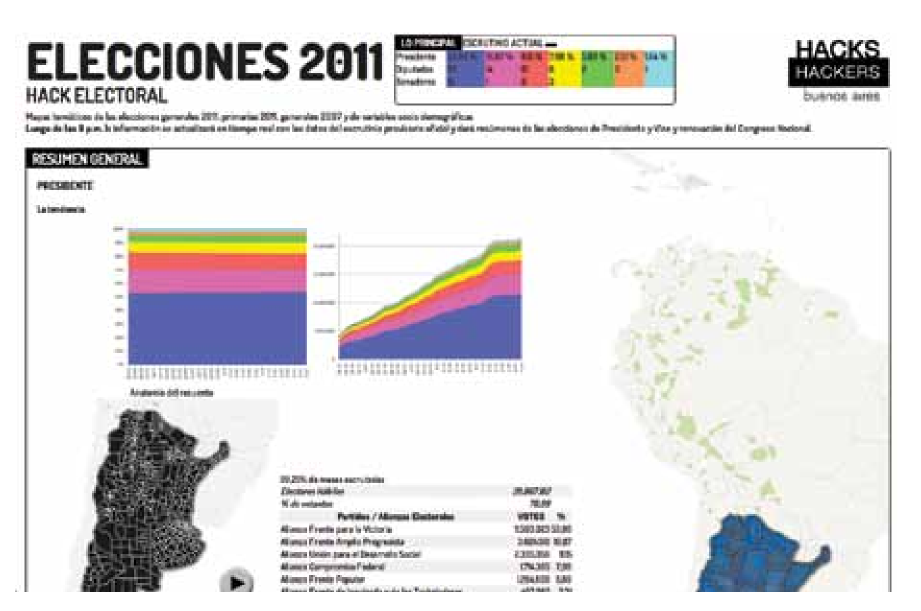
بالطبع يقوم مكتب المراجعة الوطني بنشر البيانات بالفعل، لذا لم تكن هناك حاجة إلى إعادة نشرها. ولكن بعد أن قمنا بتنقيح المعلومات وهيكلتها بشكل أفضل، قررنا نشرها. نحن نوفر بياناتنا مصحوبة برخصة إسناد من منظمة المشاع الإبداعي (http://creativecommons.org/licenses/by/3.0/). وعقب ذلك، قام العديد من المطوّرين المستقلين بإعداد توضيحات مرئية بناءً على البيانات التي قمنا بنشر البعض منها.

الأدوات التي استخدمناها في المشروع هي إكسل وجوجل ريفاين (Google Refine) لتنقية البيانات وتحليلها، وسايتوسكيب (Cytoscape) للتحليل الشبكي، وبرنامجي إلستريتور (Illustrator) وفلاش (Flash) للتوضيحات المرئية. كان من اللازم أن يكون برنامج فلاش في صيغة HTML5 ولكن الوقت لم يسعفنا.

ماذا تعلمنا؟ ربما كان الدرس الأهم هو أنّ هياكل البيانات قد تكون صعبة للغاية أحياناً. فإذا لم تكن البيانات الأصلية في صيغة مناسبة، ستستغرق عملية إعادة حسابها وتحويلها وقتاً طويلاً.

تحليل الانتخابات في الزمن الحقيقي (ملتقى الصحافيين والمبرمجين في بوينوس آيرس)

يعرف مشروع «تحليل الانتخابات» (http://elecciones.hhba.info) بأنه مشروع تحليل سياسي يقوم بإعداد توضيحات مرئية لبيانات النتائج الأولية للانتخابات التي جرت في أكتوبر 2011 في الأرجنتين. كما يعرض النظام كذلك معلومات من الانتخابات السابقة ومن الإحصاءات الديمغرافية والاجتماعية من كافة أنحاء البلاد. تم تحديث المشروع في الزمن الحقيقي بمعلومات أخذت من النتائج الأولية لصناديق الاقتراع في الانتخابات العامة الأرجنتينية لعام 2011، حيث قدمت ملخصات لنتائج الانتخابات. كانت البادرة من ملتقى الصحافيين والمبرمجين في بوينوس آيرس بالتعاون مع المحلل السياسي أندي تو، كما كانت جهداً تعاون في بذله صحافيون ومطوّرون ومصممون ومحللون وعلماء سياسيون



نموذج 3–13. انتخابات عام 2011 (ملتقى الصحافيين والمبرمجين في بوينوس آيرس)

وآخرون محليون.

ما هي البيانات التي استخدمناها؟

أتت كل البيانات من مصادر رسمية: أتاح المكتب القومي للانتخابات الوصول إلى بيانات النتائج الأولية من شركة إندرا (Indra)، كما قامت وزارة الداخلية بتوفير المعلومات حول المناصب المنتخبة والمرشحين عن مختلف الأحزاب السياسية. وقام مشروع جامعي (http: / /yoquierosaber.org/) بتوفير معلومات حول السير الذاتية والسياسات لكل مترشح للرئاسة ونائبه، بينما أتت المعلومات الديمغرافية والاجتماعية من الإحصاء القومي للسكان والإسكان (IN- DEC) لعامي 2001 و2010 ووزارة الصحة.

كيف تم تطويرها؟

تم إعداد التطبيق أثناء ملتقى الصحافيين والمبرمجين في بوينوس آيرس لتغطية انتخابات عام 2011 في اليوم الذي سبق الانتخابات: 23 أكتوبر 2011. شارك في الملتقى 30 متطوعاً من خلفيات متنوعة. تم تطوير مشروع "تحليل الانتخابات"

كمنصة مفتوحة يمكن تحسينها مع مرور الوقت. أما بالنسبة للجانب التقني، فقد استخدمنا جداول جوجل فيوجن وخرائط جوجل ومكتبات للرسوم البيانية.

قمنا بالعمل على إنشاء مضلعات لعرض الخرائط الجغرافية والتركيبة الديمغرافية. وبعد دمج المضلعات في برنامج GIS وهندسيات الجداول العامة في جوجل فيوجن، قمنا بإنشاء جداول ذات مفاتيح مقابِلة لقاعدة البيانات الانتخابية التابعة لوزارة الداخلية وشركة إندرا والبيانات الديمغرافية والاجتماعية من الإحصاء القومي للسكان والإسكان. ومن ثم، قمنا بإعداد التوضيحات المرئية باستخدام خرائط جوجل.

وباستخدام وصلة تطبيقات المستخدم ’انترفيس‘ API من خرائط جوجل، نشرنا العديد من الخرائط حسب الموضوع وتوضح التوزيع الجغرافي للأصوات بألوان متدرجة حيث تمثل كثافة اللون النسبة المئوية للأصوات لمختلف المرشحين الرئاسيين في الإدارات ومراكز الاقتراع المختلفة. وكان التركيز على المراكز الحضرية الرئيسية: مدينة بوينوس آيرس والدوائر الأربع والعشرين داخل بوينوس آيرس الكبرى ومدينتي قرطبة وروساريو.

استخدمنا نفس الوسيلة لرسم خرائط للانتخابات السابقة حسب الموضوع، وبالتحديد الانتخابات الرئاسية الأولية لعام 2011 وانتخابات عام 2007، وكذلك لتوزيع البيانات الديمغرافية والاجتماعية كتلك المتعلقة بالفقر ونسبة وفيات الأطفال وظروف المعيشة، مما يتيح فرصة للتحليل والمقارنة. كما أوضح المشروع كذلك التوزيع الجغرافي للاختلافات بين النسب المئوية للأصوات التي حصل عليها كل مرشح رئاسي ونائبه خلال الانتخابات العامة في أكتوبر الماضي، مقارنة بانتخابات أغسطس الرئيسية.

وبعد استخدامنا لبيانات جزئية من نتائج الانتخابات الأولية، قمنا برسم خريطة متحركة توضح تلك النتائج بالتفصيل، حيث تظهر حركة إحصاء الأصوات منذ غلق مراكز الاقتراع المحلية حتى صباح اليوم التالي.

الإيجابيات:

- شرعنا في العمل لإيجاد البيانات وعرضها ونجحنا في مسعانا. وقد كنا نستخدم

قاعدة بيانات اليونيسيف الخاصة بالتركيبة الديمغرافية والاجتماعية للطفل (/http://infoargentina.unicef.org.ar). كما استخدمنا قاعدة بيانات المرشحين الرئاسيين التي أعدتها مجموعة موقع yoquierosaber.org من جامعة توركواتو دي تيلا الأرجنتينية. وقد تمكنا أثناء الملتقى من جمع كم هائل من البيانات الإضافية التي لم نقم بتضمينها في النهاية.

- كان من الواضح أن عمل الصحافة والبرمجة قد تم دعمه عن طريق المنح. فبدون مساهمات آندي تو وهيلاريو مورينو كامبوس، كان من المستحيل تنفيذ هذا المشروع.

السلبيات:

- لم تكن البيانات الديمغرافية والاجتماعية المتاحة لنا حديثة أو مُفصّلة، فقد كان معظمها مأخوذاً من الإحصاء السكاني لعام 2001. فعلى سبيل المثال، لم تشمل البيانات تفاصيل حول متوسط إجمالي الناتج المحلي أو النشاط الاقتصادي الأساسي أو مستوى التعليم أو عدد المدارس أو عدد الأطباء مقابل كل فرد أو الكثير من المعلومات الأخرى التي كنا نأمل الحصول عليها.
- كان الغرض الأصلي من هذا النظام أن يعمل كأداة لدمج أي بيانات عشوائية وعرضها، وذلك حتى يتمكن الصحافيون من عرض البيانات التي تهمهم على الويب بسهولة. ولكن كان علينا تأجيل هذا الأمر لوقت لاحق.
- ولأن المشروع كان قد أنشئ بواسطة المتطوعين في فترة زمنية محدودة، كان من المستحيل أن ننفذ كل ما كنا نطمح في تنفيذه. ورغم ذلك، فقد حققنا الكثير من الإنجازات في الاتجاه الصحيح.
- ولنفس السبب، فإن تعاون ثلاثين شخصاً في العمل تم تركيزه في مبرمج واحد عندما بدأت البيانات الحكومية في الظهور، وصادفتنا بعض المشاكل في الحصول على البيانات بشكل فوري. وقد تم إيجاد الحلول خلال ساعات معدودة.

التداعيات:

كان لمشروع "تحليل الانتخابات" تأثير إعلامي كبير من خلال تغطية التلفزيون والإذاعة والصحافة المطبوعة والإلكترونية. وقد قامت جهات إعلامية عديدة

باستخدام خرائط المشروع أثناء الانتخابات وفي الأيام التي تلتها. بمرور الوقت، تم تحديث الخرائط والتوضيحات المرئية، مما زاد من نسب المشاهدة. وفي يوم الانتخابات، فإن الموقع الذي تم إنشاؤه في ذلك اليوم تحديداً استقبل حوالي عشرين ألف زائر، كما تمت إعادة نشر الخرائط على الصفحة الأولى لصحيفة "باخينا 12" ليومين على التوالي، وكذلك في مقالات بصحيفة "لا ناسيون". كما ظهرت بعض الخرائط في النسخة المطبوعة من صحيفة كلارين.

كانت تلك هي المرة الأولى التي يُستخدم فيها عرض تفاعلي لخرائط الزمن الحقيقي في تاريخ الصحافة الأرجنتينية. وقد كان فوز كريستينا فيرنانديز دي كيرشنر بنسبة 54 بالمائة من الأصوات واضحاً في الخرائط المركزية، وتم تحليله بكثافة الألوان. كما ساعد المستخدمين على فهم حالات معينة حيث حقق المرشحون المحليون في المقاطعات فوزاً ساحقاً.

— ماريانو بليجمان، ماريانا بيرويثو، سيرجيو سورين، آندي تو، مارتن سارسيل، من ملتقى الصحافيين والمبرمجين في بوينوس آيرس

البيانات في الأخبار: ويكيليكس

بدأ الموضوع بطرح إحدى فرق الصحافة الاستقصائية السؤال التالي: «أنتم تجيدون استخدام الجداول البيانية، أليس كذلك؟» وقد كان هذا جدولاً بيانياً هائلاً، إذ شمل 92,201 صفاً من البيانات، يحتوى كل منها على تحليل مفصل لحدث عسكري في أفغانستان. وكان كل ذلك الجزء الأول من سجلات الحرب من ويكيليكس (http://bit.ly/guardian-warlogs). كما أنه كان هناك جزأن لاحقان: العراق والبرقيات. كان المصطلح الرسمي هو: (SIGACT)، أي: قاعدة بيانات أهم أعمال الجيش الأمريكي.

كانت سجلات حرب أفغانستان – بمشاركتها مع صحيفتي نيويورك تايمز ودير شبيغل – هي التطبيق الفعلي لصحافة البيانات. كنا نسعى لتمكين فريقنا المكون من الصحافيين المتخصصين من الحصول على الأخبار ذات الطابع الإنساني من

خلال المعلومات المتاحة وتحليلها للوصول إلى الصورة الكاملة حتى نوضح المسار الحقيقي لتلك الحرب.

كان مهماً في تلك المرحلة المبكرة ألّا ننشر قاعدة البيانات بأكملها، فموقع ويكيليكس كان سيقوم بذلك وأردنا أن نضمن عدم الكشف عن أسماء المرشدين، كما أردنا أن نحرص على ألّا نعرض قوات حلف الناتو للخطر بلا داع. وفي الوقت ذاته، كان علينا أن نجعل استخدام البيانات أسهل لفريق الصحافيين الاستقصائيين الخاص بنا، والذي كان يقوده ديفيد لاي ونك دايفيس (هما اللذان توليا التفاوض مع جوليان أسانج على نشر البيانات). كما أردنا تبسيط الوصول إلى المعلومات الأساسية في العالم الحقيقي بشكل واضح ومتاح قدر المستطاع.

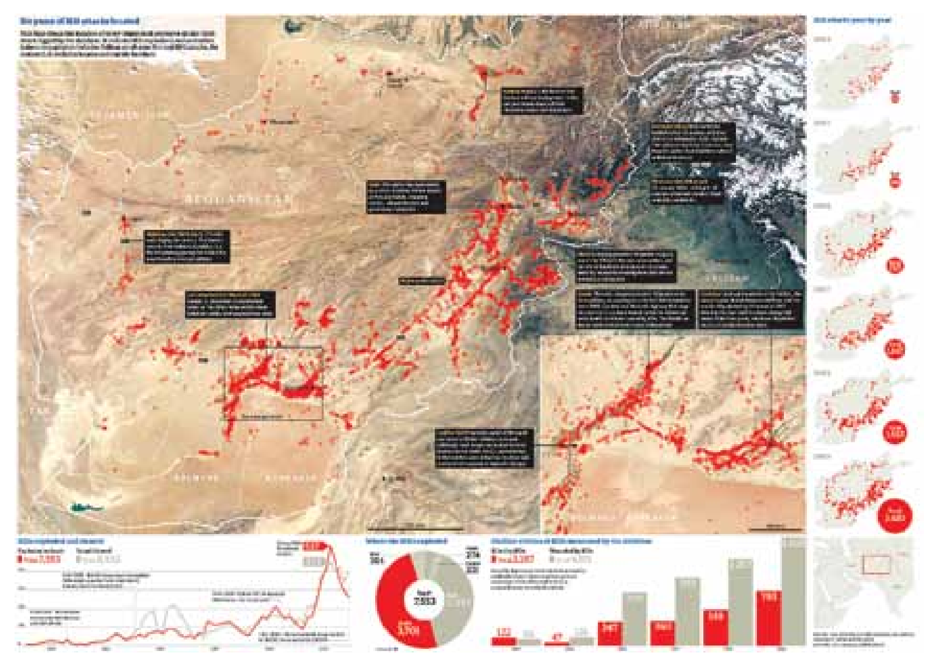
وصلتنا البيانات في ملف إكسل ضخم وشملت أكثر من 92,201 صف من البيانات، بعضها لا يحوي شيئاً على الإطلاق أو في صيغة سيئة. كما أنها لم تتح للصحافيين الخوض في البيانات للحصول على الأخبار، علاوة على حجمها الضخم الذي شكّل عائقاً أمام إعداد تقارير إخبارية ذات معنى.

أنشأ فريقنا قاعدة بيانات داخلية بسيطة باستخدام SQL، مما مكّن الصحافيين من البحث في الأخبار عن طريق كلمات بحثية أو حسب الأحداث. ثم، وبشكل مفاجئ، أصبح الوصول إلى قاعدة البيانات واستخراج الأخبار منها أسهل.

كانت البيانات منظمة بشكل جيد، إذ شمل كل حدث البيانات الرئيسية التالية: الوقت والتاريخ ووصف الحدث وعدد الضحايا وخاصة تفاصيل خطوط الطول والعرض.

كما بدأنا في فرز البيانات لتساعدنا في سرد أحد أهم أخبار الحرب: ارتفاع معدل الهجمات بواسطة العبوات الناسفة (IED) وهي قنابل يدوية الصنع لا يكمن التنبؤ بوجودها وتصعب مكافحتها. كانت مجموعة البيانات تلك لا تزال ضخمة ولكن كان التعامل معها أسهل. كان هناك حوالي 7,500 تفجير أو كمين (حيث يصاحب الهجوم إطلاق نار بواسطة الأسلحة الصغيرة أو بالقذائف) ما بين عامي 2004 و2009. كما تم العثور على 8,000 عبوة ناسفة أخرى وإزالتها. أردنا أن نقارن ونرى كيف تغير الوضع مع مرور الوقت. وقد مكنتنا البيانات من إدراك أن جنوب أفغانستان – حيث تمركزت القوات البريطانية والكندية في ذلك الوقت

– كان من أشد المناطق تضرراً من تلك الهجمات، مما أكّد معلومات صحافيينا الذين قاموا بتغطية الحرب.



نموذج 3–14. سجلات الحرب من موقع ويكيليكس (الغارديان)

وعندما نُشرت سجلات حرب العراق في أكتوبر عام 2010، تمت إتاحة 391,000 سجل إضافي لحرب العراق للجمهور.

كان هذا أمراً مختلفاً عن التسريبات الخاصة بأفغانستان، ويمكن القول إن هذا جعلها الحرب الأكثر توثيقاً في التاريخ. فقد أُتيح لنا فحص وتحليل كافة التفاصيل الدقيقة. ولكن عاملاً واحداً ظل بارزاً، وهو عدد القتلى الضخم الذين كان معظمهم من المدنيين.

وكما هو الحال في أفغانستان، فقد قررت صحيفة الغارديان ألّا تعيد نشر قاعدة البيانات بأكملها، إذ لم نكن متأكدين من احتواء خانة الملخصات على تفاصيل سرية عن المرشدين أو ما إلى ذلك.

ولكننا سمحنا للمستخدمين بتحميل جدول بياني يحتوي على سجلات كل حادثة مات خلالها شخص ما، والمجموع يقرب من 60,000 قتيل. كما قمنا بإزالة

خانة الملخصات لتبقى فقط البيانات الأساسية وهي الرتب العسكرية وعدد القتلى والتوزيع الجغرافي.

وقد وضعنا كل حوادث القتل في خريطة باستخدام جداول جوجل فيوجن (http://bit.ly/guardian-iraq-map). وعلى الرغم من أنها لم تكن مثالية، إلا أنها كانت البداية لمحاولة رصد أنماط الدمار الذي أصاب العراق.

وفي شهر ديسمبر2010 تم نشر البرقيات وهو ما اعتبر أمراً مختلفاً تماماً، فقد شملت مجموعة بيانات ضخمة تحتوي على وثائق رسمية: 251,287 رسالة من أكثر من 250 سفارة وقنصلية أمريكية في مختلف أنحاء العالم. إنها صورة فريدة للّغة الدبلوماسية الأمريكية والتي تشمل أكثر من 50,000 وثيقة تغطي إدارة أوباما الحالية. ولكن ما هو مضمون البيانات؟

وصلت البرقيات عبر شبكة الإنترنت السرية الضخمة SIPRNet، أي نظام الإنترنت العالمي الخاص بالجيش الأمريكي، وهو منفصل عن خدمة الإنترنت العامة للمدنيين وتديره وزارة الدفاع الأمريكية في واشنطن.

ومنذ هجمات سبتمبر عام 2001، بدأ التحرك في الولايات المتحدة لربط أرشيف المعلومات الحكومية لتلافي عزل معلومات الاستخبارات الأساسية في مخازن المعلومات. وقد تزايد عدد السفارات الأمريكية المرتبطة بشبكة SIPRNet خلال العقد الماضي، مما جعل تعميم المعلومات العسكرية والدبلوماسية أمراً ممكناً. وبحلول عام 2002، كانت هناك 125 سفارة أمريكية متصلة بشبكة SIPRNet، وارتفع الرقم إلى 180 بحلول عام 2005. وفي الوقت الراهن، فإن الغالبية العظمى من البعثات الأمريكية بأنحاء العالم أصبحت متصلة بهذه الشبكة. وهو ما يفسر لماذا جاء الجزء الأكبر من البرقيات خلال بين عامي 2008 و2009. في هذا الصدد كتب ديفيد لاي:

"كل رسائل السفارت تحت رمز SIPDIS يتم تحميلها تلقائياً إلى الموقع السري للسفارة المعنية. ومن هناك، يمكن الوصول إليها ليس فقط بواسطة أي شخص في وزارة الخارجية، ولكن أيضاً بواسطة أي شخص في الجيش الأمريكي ممن يملكون تصريحاً أمنياً بمستوى "سري" وكلمة سر وجهاز كومبيوتر متصل بشبكة SIPRNet."

ومما يثير الدهشة أن هذه الشبكة تغطي حوالي 3 ملايين شخص. وهناك عدة طبقات من البيانات وصولاً إلى البيانات السرية، بمعنى أنها مصممة بشكل لا يتيح إظهارها لغير مواطني الولايات المتحدة الأمريكية. وبدلاً من ذلك، من المفترض أن يطلع عليها كبار المسئولين في واشنطن وصولاً إلى مستوى وزيرة الخارجية هيلاري كلينتون. عادة يتولى صياغة البرقيات السفراء المحليون أو الموظفون التابعون لهم. أما الوثائق الاستخباراتية تحت بند "سري للغاية" أو أعلى من ذلك فلا يمكن الوصول إليها عن طريق شبكة SIPRNet.

وعلى عكس الإصدارات السابقة، فقد كان معظم تلك البيانات عبارة عن نص مكتوب غير محدد كمّياً ولا يتضمن بيانات متطابقة. وقد شملت المحتويات ما يلي:

المصدر: أي السفارة أو الجهة المرسلة.

قائمة المتلقين: عادة ما توجه البرقيات إلى عدد من السفارات والجهات الأخرى.

موضوع الرسالة: أي ملخص البرقية.

كلمات دليلية: تم إدراج كلمات دليلية في كل برقية تحمل عدداً من الاختصارات الأساسية.

نص الرسالة: أي مضمون البرقية ذاتها. فضّلنا عدم نشر النص الكامل للرسائل لأسباب أمنية واضحة.

إحدى الملاحظات المثيرة للاهتمام هنا أن تلك البرقيات قد وفرت ما يمكن وصفه تقريبا "تسريبات عند الطلب"، إذ إنها ظلت تتصدر عناوين الأخبار لمدة أسابيع عقب نشرها. والآن كلما نُشر خبر عن نظام فاسد أو عن فضيحة دولية، فإن الوصول إلى تلك البرقيات يتيح لنا الحصول على أخبار جديدة.

يعتبر تحليل البرقيات عملية شاقة قد لا يمكن إنهاؤها بالكامل أبداً.

—— هذا النص نسخة منقحة من فصل تم نشره للمرة الأولى في كتاب "الحقائق مقدسة: قوة البيانات" لمؤلفه سايمون روجرز، صحيفة الغارديان (نُشر على كيندل).

ملتقى Mapa76 للمبرمجين

قمنا بفتح باب ملتقى الصحافيين والمبرمجين في بوينوس آيرس (http://www.meetup.com/HacksHackersBA/) في أبريل عام 2011. ثم استضفنا لقاءين أوليين للترويج لفكرة التعاون الواسع بين الصحافيين ومطوري البرامج، بحضور ما بين 120 و150 شخصاً في كل لقاء. وفي الاجتماع الثالث نظمنا ملتقى للصحافيين والمبرمجين دام 30 ساعة مع تكليف ثمانية أشخاص بحضور مؤتمر الصحافة الرقمية بمدينة روزاريو الواقعة على بعد 300 كيلومتر من بوينوس آيرس.

كان الموضوع السائد في هذه الاجتماعات الرغبة في استخراج كم هائل من البيانات من الويب ثم عرضها مرئياً. وللمساعدة في إنجاز هذا العمل، تم إنشاء مشروع يُدعى Mapa76.info ليساعد المستخدمين على استخراج البيانات ثم عرضها من خلال الخرائط والترتيب الزمني. ولم تكن تلك مهمة سهلة.



نموذج 3–15. ملتقى Mapa76 للمبرمجين
(ملتقى الصحافيين والمبرمجين في بوينوس آيرس)

ما الهدف من مشروع Mapa76؟ في 24 مارس 1976 حدث انقلاب عسكري في الأرجنتين واستمر حكم قادته حتى عام 1983. في ذلك الوقت، قُدر عدد المفقودين بـ 30 ألفاً، بالإضافة إلى آلاف القتلى، وولادة 500 طفل داخل السجون لمن سُجنَّ بأوامر من السلطة الدكتاتورية العسكرية. بعد مرور أكثر من 30 عاماً، بلغ عدد الذين تمت إدانتهم بارتكاب جرائم ضد الإنسانية خلال الحكم الدكتاتوري

262 شخصاً (حتى سبتمبر 2011). هناك حالياً 14 محاكمة جارية وسبع محاكمات تم تحديد تواريخ بدئها. هناك أيضاً 802 شخصاً يخضعون لمحاكمات مفتوحة في عدة قضايا.

ينتج عن هذه المحاكمات كم هائل من البيانات يصعب على الباحثين والصحفيين ومنظمات حقوق الإنسان والقضاة ومسؤولي الادعاء العام وغيرهم التعامل معه. يتم إنتاج وتوزيع البيانات، ولكن غالباً ما لا يستفيد المحققون من البرامج لمساعدتهم في تفسيرها. وهذا حتماً يعنى أنه غالباً ما يتم التغاضي عن الحقائق، وعادة ما تكون الفرضيات محدودة. يُعد مشروع Mapa76 بمثابة أداة استقصائية تتيح الوصول إلى هذه المعلومات لأغراض صحافية وقانونية وقضائية وتاريخية.

من أجل التحضير للملتقى، أنشأنا منصة يمكن للمطورين والصحافيين الاستفادة منها للتعاون فيما بينهم أثناء الفعالية. وقد قام مارتن سارسيل بتطوير بعض الحسابات الأساسية لاستخراج البيانات المنظمة من الوثائق النصية البسيطة. كما تم استخدام بعض المكتبات من مشروع DocumentCloud.org ولكن ليس بشكل كبير. تقوم المنصة تلقائياً بتحليل واستخراج الأسماء والتواريخ والأماكن من النصوص مما يتيح للمستخدمين استكشاف الحقائق الأساسية عن مختلف القضايا، كتاريخ الميلاد ومكان الاعتقال ومكان الاختفاء المزعوم وما إلى ذلك.

كان هدفنا توفير منصة لاستخراج البيانات تلقائياً حول الأحكام الصادرة من النظام العسكري الدكتاتوري في الأرجنتين. وكنا نود الوصول إلى طريقة تلقائية (أو على الأقل شبه تلقائية) لعرض بيانات أساسية تتعلق بقضايا خلال الفترة من 1976 إلى 1983 استناداً إلى أدلة مكتوبة ومدعومة بالحجج والأحكام. يتم جمع البينات المستخرجة (الأسماء والأماكن والتواريخ) وتخزينها وتحليلها وفرزها بواسطة الباحثين وكذلك استكشافها باستخدام الخرائط والترتيب الزمني وأدوات التحليل الشبكي.

سيتيح المشروع للصحافيين والمحققين ومسؤولي الادعاء العام والشهود متابعة قصص حياة الأشخاص بما في ذلك أحوالهم داخل السجون وما يعقب ذلك من اختفائهم أو إطلاق سراحهم. وفي حالة غياب المعلومات، يمكن للمستخدمين

البحث في كم هائل من الوثائق للعثور على المعلومات التي قد تكون ذات صلة ببعض القضايا.

قمنا بالإعلان عن الملتقى من خلال منصة الصحافيين والمبرمجين في بوينوس آيرس (/http://www.meetup.com/HacksHackersBA) التي بلغ عدد أعضائها حوالي 200 شخص (وقد وصل العدد إلى 540 عضواً أثناء كتابة هذا الكتاب). كما أننا اتصلنا بالعديد من منظمات حقوق الإنسان. حضر الاجتماع حوالي 40 شخصاً، بما فيهم صحافيون ومنظمات للمناصرة ومطوّرون ومصممون.

في أثناء الملتقى، قمنا بتحديد المهام التي يمكن لمختلف فئات المشاركين تأديتها بمفردهم للمساعدة في إنجاز الأعمال بسلاسة. على سبيل المثال، كلفنا المصممين بإعداد "إنترفيس" للتطبيق يجمع بين الخرائط والتتابع الزمني للأحداث. كما طلبنا من المطوّرين البحث عن وسائل لاستخراج البيانات المنظمة والحسابات لإزالة الغموض عن الأسماء. وطلبنا أيضاً من الصحافيين تحري ما حدث لأشخاص معينين بهدف مقارنة النسخ المختلفة للخبر والبحث في الوثائق لسرد أخبار تتعلق بقضايا معينة.

ربما كانت المشكلة الرئيسية التي واجهتنا بعد الملتقى هي أن مشروعنا كان طموحاً أكثر مما ينبغي، كما أن أهدافنا القصيرة المدى كانت تتطلب الكثير من الجهد والوقت. من الصعب التنسيق بين شبكة من المتطوعين غير المترابطين. كان كل من يعمل على المشروع تقريباً مشغولاً بعمله طوال اليوم، كما شارك العديد منهم في فعاليات ومشاريع أخرى، إذ تم عقد تسعة اجتماعات أثناء الملتقى عام 2011.

يخضع المشروع حالياً للتطوير النشط. وهناك فريق أساسي مكون من أربعة أشخاص يعملون مع أكثر من اثني عشر متعاونا. كما أن لدينا قائمة بريدية عامة (/http://groups.google.com/group/mapa76-dev) ومركزاً للترميز "الكود" (/http://github.com/mapa76) حيث يمكن لأي شخص الاشتراك في المشروع.

— ماريانو بليغمان، ملتقى الصحافيين والمبرمجين في بوينوس آيرس

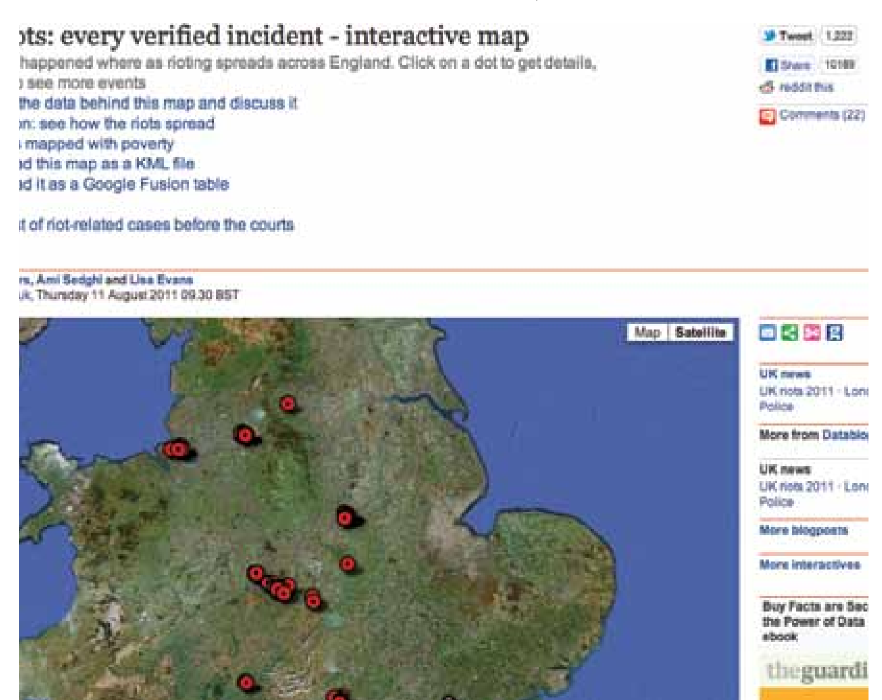
تغطية مدونة البيانات بصحيفة الغارديان لأحداث الشغب في المملكة المتحدة

في صيف عام 2011 تعرضت بريطانيا لموجة من أحداث الشغب. وقد قال السياسيون حينها إن تلك الأعمال لا صلة لها أبداً بالفقر وإن من قاموا بالسلب والنهب مجرد مجرمين. كما قام رئيس الوزراء وكبار الساسة المحافظين بإلقاء اللوم على الإعلام الاجتماعي لتسببه في أحداث الشغب، قائلين إنه قد تم استخدام تلك المنصات من أجل التحريض وإن أحداث الشغب قد تم تنظيمها من خلال فيسبوك وتويتر ورسائل بلاكبيري (BBM). كانت هناك دعوات لإغلاق الإعلام الاجتماعي بشكل مؤقت. ولأن الحكومة لم تجرِ تحقيقاً في سبب اندلاع أحداث الشغب، تعاونت صحيفة الغارديان مع كلية لندن للاقتصاد وقاما معاً بإنشاء المشروع الرائد 'قراءة أحداث الشغب' 'Reading The Riots' (http://www.guardian.co.uk/uk/series/reading-the-riots) للتطرق إلى تلك القضايا.

استخدمت الصحيفة صحافة البيانات بشكل مكثف لتمكين الناس من فهم من يقوم بالسلب والنهب، وما السبب في ذلك. كما أنهم عملوا مع فريق آخر من الأكاديميين، بقيادة الأستاذ روب بروكتر من جامعة مانشستر، لفهم دور الإعلام الاجتماعي بشكل أفضل، والذي استخدمته صحيفة الغارديان نفسها بشكل مكثف في إعداد تقاريرها أثناء أحداث الشغب. وقد ترأس فريق 'قراءة أحداث الشغب' بول لويس، محرر المشاريع الخاصة لدى صحيفة الغارديان. أثناء الأحداث، نقل بول الأخبار من الجبهة في عدة مدن بريطانية (ولا سيما من خلال حسابه على تويتر، @paullewis). أما الفريق الثاني فقد عمل على 2.6 مليون تغريدة عن الأحداث ساهم بها تويتر. وكان الهدف الرئيسي لهذا العمل على الإعلام الاجتماعي رصد كيفية انتشار الشائعات على تويتر، ودور مختلِف المستخدمين أو الفاعلين في الترويج لتدفق المعلومات ونشرها، ورؤية ما إذا كانت المنصة قد استخدمت للتحريض، وفحص أشكال أخرى من التنظيم.

فيما يتعلق باستخدام صحافة البيانات وتوضيح البيانات مرئياً، فمن المفيد أن نميز بين فترتين أساسيتين: فترة الأحداث نفسها وطرق سرد البيانات للأخبار مع

تطور أحداث الشغب، ثم الفترة الأخرى من البحث المكثف مع مجموعتين من الفرق الأكاديمية التي تعمل مع الغارديان لجمع البيانات وتحليلها وكتابة تقارير متعمقة حول نتائج البحث. تم نشر نتائج المرحلة الأولى لمشروع 'قراءة أحداث الشغب' خلال أسبوع من التغطية واسعة النطاق في ديسمبر 2011. فيما يلي بعض الأمثلة الأساسية لكيفية استخدام صحافة البيانات خلال كلتا الفترتين.



نموذج 3–16. أحداث الشغب في المملكة المتحدة: كل الحوادث المؤكدة: خريطة تفاعلية (الغارديان)

المرحلة الأولى: أحداث الشغب كما حدثت بالفعل

عن طريق استخدام الخرائط، عرض فريق البيانات لدى صحيفة الغارديان المواقع المؤكدة لأحداث الشغب (http://bit.ly/guardian-riots-map)، ومن خلال مزج بيانات الحرمان الاجتماعي مع مواقع أحداث الشغب (http://bit.ly/guardian-riots-poverty)، بدأ الفريق في فضح الرواية السياسية

الأساسية التي تفيد بأن الأحداث لا تتعلق بالفقر. وقد استخدم كلا المثالين أدوات تخطيط جاهزة، وفي المثال الثاني، جمع بين بيانات الموقع ومجموعة بيانات أخرى للبدء في صنع صلات وروابط أخرى.

فيما يتعلق باستخدام الإعلام الاجتماعي أثناء الأحداث، تويتر في هذه الحالة، وضعت الصحيفة توضيحاً مرئياً لعلامات الهاشتاج # المستخدمة أثناء تلك الفترة (http://bit.ly/guardian-riots-twitter)، مما سلط الضوء على أن تويتر قد استخدم أساساً لتنسيق رد فعل لأحداث الشغب وليس لتنظيم الناس للقيام بالسلب والنهب. وقد كان هاشتاج #riotcleanup، وهو للحملة العفوية لتنظيف الشوارع بعد أحداث الشغب، هو الأكثر استخداماً أثناء الأحداث.

المرحلة الثانية: قراءة أحداث الشغب

عندما نشرت الصحيفة النتائج بعد بحث مكثف دام عدة أشهر وبعد العمل عن قرب مع فريقين أكاديميين، برز توضيحان مرئيان وتمت مناقشتهما على نطاق واسع. أولهما كان مقطع فيديو قصيراً (http://bit.ly/guardian-riots-commute) يعرض نتائج جميع الأماكن التي قام الناس بأحداث الشغب فيها مع عناوين منازلهم، ويعرض الفيديو ما يسمى بـ "الانتقال إلى أحداث الشغب". وهنا عملت الصحيفة مع المتخصصين في رسم خرائط النقل، من شركة ITO World، لوضع نموذج لأكثر المسارات احتمالاً لاستخدام مثيري الشغب في طريقهم إلى مواقع مختلفة للقيام بأعمال السلب والنهب، كما سلطوا الضوء على أنماط مختلفة لمدن مختلفة، حيث قطع البعض مسافات طويلة.

أما التوضيح المرئي الثاني فيتطرق إلى كيفية انتشار الشائعات على تويتر. في نقاش مع الفريق الأكاديمي، تم الاتفاق على سبع شائعات من أجل التحليل. قام بعد ذلك الفريق الأكاديمي بجمع كل البيانات المتعلقة بكل شائعة وابتكر جدولاً لترميز التغريدة وفقاً لأربعة رموز أساسية: الذين يكررون الشائعة فقط (الادعاء)، والذين يرفضونها (الادعاء المضاد)، والذين يشككون فيها (التساؤل)، والذين يعلقون عليها ببساطة (التعليق). تم ترميز جميع التغريدات في ثلاث نسخ، كما قام فريق الغارديان التفاعلي بتوضيح النتائج مرئياً (-http://bit.ly/guardian

riots). ثم كتب الفريق موضحاً كيفية بناء التوضيح المرئي (/http://bit.ly
.(guardian-riots-twitter-interactive

الأمر المذهل بشأن هذا التوضيح المرئي هو أنه يظهر بوضوح صارخ ما يصعب جداً وصفه وهو سرعة انتشار الشائعات وتطور دورة حياتها مع مرور الوقت. يتضح دور وسائل الإعلام الرئيسية السائدة في بعض من تلك الشائعات (فضح زيفها بصراحة أو تأكيدها سريعا كأخبار مثلاً)، مثل الطبيعة التصحيحية لتويتر نفسه فيما يتعلق بالتعامل مع مثل تلك الشائعات. لم يقتصر دور هذا التوضيح المرئي على المساعدة في سرد الخبر بشكل كبير فحسب، بل أيضاً أعطى رؤية حقيقية لكيفية عمل الشائعات على تويتر، مما وفر معلومات مفيدة لكيفية التعامل مع الأحداث المماثلة في المستقبل.

يتضح لنا من المثال السابق التعاون القوي بين الصحيفة والفريق الأكاديمي الذي استطاع إجراء تحليل متعمق لـ 2.6 مليون تغريدة عن أحداث الشغب. وعلى الرغم من أن الفريق الأكاديمي قام ببناء مجموعة من الأدوات حسب الطلب ليتمكن من إجراء التحليل، إلا أنهم يعملون الآن على إتاحتها بشكل واسع لمن يريد استخدامها (/http://www.analysingsocialmedia.org) في الوقت المناسب، مما يوفر نمط عمل للتحليل. وبجمع تلك الأدوات مع الشرح الذي يوفره فريق الغارديان، ستتوفر دراسة حالة مفيدة لكيفية استخدام الآخرين لتحليلات الإعلام الاجتماعي والتوضيحات المرئية لسرد الأخبار الهامة.

—— فريدة فيس، جامعة ليستر

الشهادات الدراسية في مدارس ولاية إلينوي

يقوم مجلس التعليم في ولاية إلينوي كل عام بنشر "الشهادات الدراسية" وبيانات التركيبة الديمغرافية والأداء لجميع المدارس الحكومية في الولاية. وهي مجموعة بيانات ضخمة، فقد بلغ عرضها هذا العام 9,500 عمود. إن مشكلة مثل هذا الكم من البيانات تكمن في اختيار ما نقدمه. (مثلما هو الحال في أي مشروع برمجي، فالجزء الصعب ليس في بناء برنامج، بل في بناء البرنامج المناسب).

عملنا مع الصحافيين والمحرر من فريق التعليم لاختيار بيانات مثيرة للاهتمام.

(هناك الكثير من البيانات التي قد تبدو مثيرة للاهتمام ولكن قد يخبرك الصحفي بأنها خطأ أو مضللة).

وقد أجرينا أيضاً دراسة ومقابلات مع من يعملون معنا في غرفة الأخبار ممن لديهم أطفال في سن الدراسة. فعلنا ذلك بسبب وجود فجوة في التعاطف – فلم يكن لدى أي من أعضاء فريق تطبيقات الأخبار أطفال في سن الدراسة. وأثناء ذلك، تعلمنا الكثير عن مستخدمينا والكثير عن قابلية استخدام النسخة السابقة (أو عدمها) من موقع المدارس الخاص بنا.



نموذج 3–17. الشهادات الدراسية في مدارس إلينوي 2011 (شيكاغو تريبيون)

كنا نهدف إلى إعداد تصميمات لمستخدمين محددين وحالات استخدام معينة تتمثل في:

- الآباء الذين لديهم طفل في المدرسة ويرغبون في معرفة مستوى مدرسته.
- الآباء الذين يحاولون أن يقرروا أين يسكنون، بما أن جودة المدرسة غالباً ما تؤثر بشكل كبير على هذا القرار.

في المرة الأولى، كان الموقع الإلكتروني للمدرسة مشروعا استغرق ستة أسابيع وعمل عليه مطوّران. أما التحديث فكان مشروعاً استغرق أربعة أسابيع وعمل عليه مطوّران. (في الحقيقة كان هناك ثلاثة أشخاص نشطين يعملون على

المشروع الأخير، ولكن لم يكن أي منهم يعمل بدوام كامل، لذا فقد اعتبرناهم شخصين).

كان تصميم المعلومات جزءاً أساسياً من هذا المشروع. وعلى الرغم من أننا نقدم بيانات أقل كثيراً مما هو متاح، إلا أنها لا تزال تُعتبر كما ضخما من البيانات، وكان التحدي أن نجعلها سهلة الفهم. لحسن الحظ، تمكنَّا من استعارة أحد العاملين لدينا في مكتب الغرافيكس، وهو مصمم متخصص في عرض المعلومات المعقدة. وقد علّمنا الكثير عن تصميم الرسوم البيانية وأرشدنا عموماً إلى عرض قابل للقراءة ولكنه لا يستهين بقدرة القارئ أو رغبته في فهم الأرقام.

تم بناء الموقع باستخدام لغتي بايثون وجانغو. وتم وضع البيانات في قاعدة بيانات MongoDB– فبيانات المدارس غير متجانسة وذات تسلسل هرمي، مما جعلها غير مناسبة لقواعد البيانات الترابطية (وإلا كنا غالباً ما سنستخدم PostgreSQL).

قمنا لأول مرة بتجربة إطار وصلة "انترفيس" المستخدم Bootstrap في موقع تويتر، ونحن راضون عن النتائج. وقد تم إعداد الرسوم البيانية بواسطة Flot.

وقد كان هذا التطبيق مصدراً للعديد من الأخبار التي كتبناها عن أداء المدارس. وبهذا يعمل التطبيق عمل البوابة نوعاً ما، فعندما يكون هناك خبر جديد عن أداء المدارس نضعه في أعلى التطبيق، إلى جانب قائمة بالمدارس ذات الصلة بالخبر. (وعندما يصدر خبر جديد، يتم وصل قراء صحيفة شيكاغو تريبيون بالتطبيق وليس بالخبر).

أفادت التقارير الأولية بأن التطبيق أعجب القراء كثيراً. وقد كانت معظم الردود التي تلقيناها إيجابية (أو على الأقل بنّاءة!)، وارتفعت زيارات الصفحة كثيراً. وكميزة إضافية، ستبقى هذه البيانات مثيرة للاهتمام لمدة عام كامل. لذلك، وعلى الرغم من أننا نتوقع انخفاض عدد الزيارات مع تلاشي أخبار المدارس من الصفحة الرئيسية، إلا أننا من واقع تجاربنا السابقة نعلم أن القراء لجأوا إلى هذا التطبيق طوال العام.

فيما يلي بعض الأفكار الأساسية التي استخلصناها من هذا المشروع:

- إن مكتب الغرافيكس صديقك. فهُم يجيدون تسهيل فهم المعلومات المعقدة وتبسيطها.

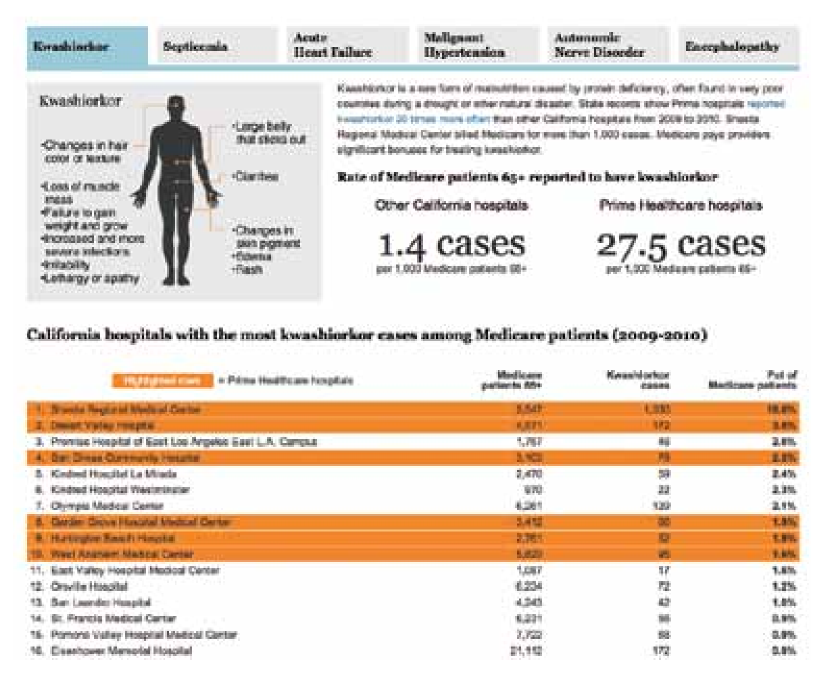
- اطلب المساعدة من غرفة الأخبار. هذا هو مشروعنا الثاني الذي نجري فيه دراسات ومقابلات في غرفة الأخبار، وهذه وسيلة جيدة للحصول على آراء ممن يفكرون، وهُم مثل جمهورنا، تتنوع خلفياتهم ولا يرتاحون عموماً إلى التعامل مع أجهزة الكمبيوتر.
- اعرض عملك! إن معظم الردود التي حصلنا عليها كانت طلبات للبيانات التي استخدمها التطبيق. وقد قمنا بإتاحة الكثير من البيانات من خلال وصلة "انترفيس" برمجة التطبيق، وسننشر قريباً الأمور التي لم نفكر في ضمها منذ البداية.

—— بريان بوير، صحيفة شيكاغة تريبيون

فواتير المستشفيات

تلقى الصحافيون الاستقصائيون في CaliforniaWatch (http://californiawatch.org/) معلومات تفيد بأن هناك سلسلة مستشفيات كبيرة في كاليفورنيا من المحتمل أنها تتلاعب بشكل منظم ببرنامج الرعاية الصحية الفيدرالي الذي يتولى تكاليف علاج الأمريكيين الذين تبلغ أعمارهم 65 عاماً أو أكثر. وقد كانت عملية الاحتيال تُدعى 'upcoding'، أي إعداد تقارير تفيد بأن حالات المرضى أكثر تعقيداً مما هي عليه بالفعل، وهكذا يحصلون على تعويضات أكثر. ولكن المصدر الرئيسي كان نقابة تحارب إدارة سلسلة المستشفيات، وقد أدرك فريق CaliforniaWatch أن تأكيد المعلومة من مصدر مستقل ضروري من أجل مصداقية الخبر.

لحسن الحظ، تحتفظ وزارة الصحة في كاليفورنيا بسجلات عامة تحتوي على معلومات مفصلة عن كل حالة تم علاجها في جميع مستشفيات الولاية. تحتوي المتغيرات الـ 128 على ما يصل إلى 25 رمزاً تشخيصياً من كتيب 'التصنيف الإحصائي الدولي للأمراض والمشاكل الصحية ذات الصلة' (والمعروف بـ ICD-9) الذي نشرته منظمة الصحة العالمية. وعلى الرغم من أن المرضى ليسوا مذكورين بالاسم في البيانات، إلا أن هناك متغيرات أخرى تشير إلى عمر المريض وكيفية دفع التكاليف وأي المستشفيات تم علاجه فيها. وقد أدرك الصحافيون أنهم



نموذج 3–18. مرض كواشيوركور (CaliforniaWatch)

يستطيعون عن طريق هذه السجلات أن يروا ما إذا كانت المستشفيات التي تملكها هذه السلسلة تبلّغ عن حالات معينة وغريبة بمعدلات أعلى كثيراً من المستشفيات الأخرى.

كانت مجموعات البيانات كبيرة، حوالي 4 ملايين سجل سنويا، وأراد الصحافيون دراسة سجلات 6 سنوات حتى يتمكنوا من رؤية كيف تتغير الأنماط مع مرور الوقت. طلبوا البيانات من وكالة الدولة، وتلقوها في صيغة أقراص مدمجة تم نسخها بسهولة إلى جهاز الكمبيوتر. استخدم الصحافي الذي كان يقوم بتحليل البيانات نظاما يُدعى SAS (http://www.sas.com/) ليعمل على البيانات. وهو نظام غاية في الكفاءة (يسمح بتحليل ملايين السجلات) وتستخدمه العديد من الوكالات الحكومية، بما فيها وزارة الصحة في كاليفورنيا، ولكنه باهظ الثمن – فقد كان من الممكن القيام بنفس التحليل باستخدام أي أدوات قواعد بيانات أخرى، مثل مايكروسوفت أكسِس Microsoft Access أو الأداة مفتوحة المصدر MySQL (http://www.mysql.com/).

بعد الحصول على البيانات وكتابة البرامج لدراستها، كان العثور على الأنماط المريبة أمراً بسيطاً نسبياً. فمثلاً، كان أحد الادعاءات يفيد بأن سلسلة المستشفيات تبلّغ عن عدة درجات من سوء التغذية بمعدلات أعلى كثيراً من المستشفيات الأخرى. باستخدام نظام SAS، استخرج محلل البيانات جداول التكرار التي أظهرت عدد حالات سوء التغذية التي يتم الإبلاغ عنها سنوياً من قِبل كلٍ من مستشفيات الرعاية الفائقة التي يزيد عددها عن الثلاثمائة في كاليفورنيا. تم بعد ذلك استيراد جداول التكرار الخام إلى إكسل للتدقيق في أنماط كل مستشفى. وقد استطاعت قدرة إكسل على الفرز والتصفية وحساب المعدلات من الأرقام الخام أن تسهل رصد الأنماط.

ومما لفت الانتباه على وجه خاص تقارير عن مرض يُدعى 'كواشيوركور'، أي متلازمة نقص البروتين التي توجد بشكل شبه حصري في الأطفال الجوعى في البلدان النامية التي تعاني من المجاعات. ومع ذلك، أبلغت سلسلة المستشفيات محل التدقيق عن تشخيص حالات كواشيوركور بين المسنين من سكان كاليفورنيا بمعدلات تصل إلى 70٪ أعلى (http://bit.ly/californiawatch-malnutrition) من المعدل المتوسط في جميع مستشفيات الولاية.

وفي الأخبار الأخرى، استخدم التحليل أساليب أخرى لفحص المعدلات التي تم الإبلاغ عنها في أمراض مثل تسمم الدم والاعتلال الدماغي وارتفاع ضغط الدم الخبيث واضطراب الجهاز العصبي الذاتي (http://bit.ly/californiawatch–rare). وقد نظر تحليل آخر في الادعاءات بأن سلسلة المستشفيات تقبل إدخال نسب مرتفعة بشكل غير عادي من مرضى التأمين الصحي الخاص 'ميديكير' المحولين من غرف الطوارئ بالمستشفى إلى أسرة العلاج بنفس المستشفى، ويتميز هؤلاء المرضى بأن لديهم مصدرا مضمونا لدفع تكاليف المستشفى مقارنة بحالات عديدة أخرى من المرضى في غرف الطوارئ.

والخلاصة أن مثل تلك الأخبار تصبح ممكنة عندما تستخدم البيانات لتصنع دليلاً يختبر– بشكل مستقل – ادعاءات المصادر التي قد تكون لديها مصالح خاصة. كما أن تلك الأخبار تعتبر مثالا جيدا على ضرورة وجود قوانين صارمة للسجلات العامة. والسبب الذي يجعل الحكومة تطلب من المستشفيات الإبلاغ عن تلك

البيانات هو أن هذا النوع من التحليلات يمكن أن تقوم به الحكومة أو الأكاديميون أو المحققون أو حتى الصحافيون المدنيون. وموضوع هذه الأخبار هام لأنه يحقق فيما إذا كانت الملايين من الدولارات من الأموال العامة تُنفق بشكل مناسب.

—— ستيف دويغ، كلية والتر كرونكايت للصحافة، جامعة أريزونا

أزمة دور الرعاية

كشف تحقيق صحيفة فاينانشال تايمز (http://on.ft.com/care-home-crisis) في صناعة دور الرعاية الخاصة عن قيام بعض المستثمرين في الأسهم الخاصة بتحويل رعاية المسنين إلى آلة ربحية، كما سلط التحقيق الضوء على التكلفة البشرية القاتلة للنموذج التجاري الذي فضّل عائدات الاستثمار على الرعاية الجيدة.

كان توقيت التحليل مناسباً لأن المشاكل المالية لدى شركة Southern Cross – وهي أكبر شركة لإدارة دور الرعاية في البلاد – كانت قد بلغت ذروتها. وعلى مدى عدة عقود، دعت الحكومة لخصخصة قطاع الرعاية وواصلت الإشادة بالذكاء التجاري للقطاع الخاص.

بدأ تحقيقنا بتحليل البيانات التي حصلنا عليها من المسؤول عن تفقد دور الرعاية في المملكة المتحدة. كانت المعلومات متاحة للجمهور، ولكنها تطلبت الكثير من المثابرة لوضع البيانات في صيغة قابلة للاستخدام.

تضمنت البيانات تقييمات (لاغية الآن) لأداء دور الرعاية وتصنيفا للدور من حيث كونها خاصة أو حكومية أو غير ربحية. وقد قامت لجنة جودة الرعاية (CQC) حتى يونيو 2010 بتقييم دور الرعاية حسب جودتها (عدد النجوم صفر = ضعيف، ثلاث نجوم = ممتاز).

تطلبت الخطوة الأولى تنقية شاملة للبيانات، ذلك أن البيانات التي وفرتها 'لجنة جودة الرعاية Care Quality Commission' احتوت مثلاً على تصنيفات غير موحدة. وقد فعلنا ذلك أولاً باستخدام إكسل. كما حددنا – من خلال البحث باستخدام الهاتف والبحث المكتبي – ما إذا كانت دور رعاية معينة تمتلكها شركات مضاربات الأسهم الخاصة. قبل الأزمة المالية، كان قطاع دور الرعاية يجذب

مستثمري الأسهم الخاصة والعقارات كالمغناطيس، ولكن بدأ العديد منها – كشركة Southern Cross – يواجه صعوبات مالية كبيرة. وقد أردنا أن نحدد تأثير ملكية الأسهم الخاصة على جودة الرعاية.

وقد مكنتنا مجموعة حسابات إكسل مباشرة نسبياً من التأكيد على أن متوسط أداء دور الرعاية الحكومية أو غير الربحية أفضل كثيراً من القطاع الخاص. وقد كان أداء بعض دور الرعاية التي تملكها شركات الأسهم الخاصة أعلى من المتوسط، بينما كان أداء دور رعاية أخرى أقل من المتوسط.

إضافة إلى الصحافة الميدانية ودراسات حالات الإهمال والفحص المتعمق في إخفاق السياسات المنظمة وبيانات أخرى عن مستويات الرواتب ومعدلات دورة تبديل الموظفين إلخ، استطاع تحليلنا أن يرسم صورة للحالة الحقيقية لدور الرعاية.

بعض النصائح:

- احرص على تدوين ملاحظات عن كيفية تغييرك للبيانات الأصلية.
- احتفظ بنسخة من البيانات الأصلية ولا تغيرها أبداً.
- افحص البيانات أكثر من مرة وقم بالتحليل عدة مرات (بالعودة إلى البداية إذا تطلب الأمر ذلك).
- إذا ذكرت شركات أو أفراداً معينين، امنحهم حق الرد.

—— سينثيا أومورشو، صحيفة فاينانشيال تايمز

الهاتف الفاضح

لدى معظم الناس معرفة نظرية عما يمكن فعله بالبيانات التي توفرها هواتفنا المحمولة؛ ولكن هناك القليل من الأمثلة الواقعية. ولهذا قرر مالتي سبيتز من حزب الخضر الألماني أن ينشر بياناته الخاصة. للحصول على تلك المعلومات، اضطر سبيتز إلى مقاضاة شركة الاتصالات العملاقة دويتشه تيليكوم Deutsche Telekom. وقد مثلت تلك البيانات – التي كانت في وثيقة إكسل ضخمة – الأساس للخريطة التفاعلية على الموقع الإلكتروني لصحيفة دي تسايت، حيث يمثل كل من الـ 35,831 صفاً في الجدول البياني كل مرة نقل فيها هاتف سبيتز معلومات على مدى نصف عام.

وإذا نظرت إلى كل جزء من البيانات على حدة، ستجد أنها في الغالب غير مؤذية. ولكن إن نظرت إلى جميع الأجزاء معاً، ستجد أنها تعطي ما يطلق عليه المحققون الملف الشخصي: صورة واضحة ونمطاً محدداً لعادات الشخص وتفضيلاته وبالطبع، حياته. يكشف هذا الملف عن الوقت الذي سار فيه سبيتز في الشارع أو عندما استقل القطار أو كان في طائرة. ويوضح أيضاً أنه يعمل أساساً في برلين ويُظهر أي المدن زارها ومتى كان مستيقظاً أو نائماً.



نموذج 3–19. الهاتف الفاضح (صحيفة دي تسايت الإلكترونية)

أبقت مجموعة بيانات شركة دويتشه تيليكوم جزءاً من سجل بيانات سبيتز خاصاً، أي الجزء الخاص باتصالاته. لأن هذا النوع من المعلومات لا ينتهك خصوصية الكثيرين في حياته فحسب، بل يكشف أيضاً عن الكثير من المعلومات بشأن سبيتز نفسه، حتى وإن كانت الأرقام مشفرة (ولكن يمكن للوكلاء الحكوميين في العالم الحقيقي أن يحصلوا على تلك المعلومات).

أسعدنا الحظ أن نعمل مع لورينز ماتزات ومايكل كريل من OpenDataCity وأن نطلب منهما أن يبحثا في البيانات ويجداً حلاً لكيفية فهم البيانات المتعلقة بالموقع الجغرافي له واستخلاصها. كان من الضروري عمل خطوط مثلثة تربط

كل اتصال تم من هاتف سبيتز مع مواقع أعمدة الهوائيات الموصلة بالطرف الآخر. ويحتوي كل صارٍ من الصواري على ثلاثة هوائيات يغطي كل منها زاوية مقدارها 120 درجة. وقد وجد المبرمجان أن الموقع المحفوظ قد أشار إلى الاتجاه الذي كان يتصل منه هاتف سبيتز المحمول.

وعند إجراء مضاهاة لمطابقة إحداثيات هذه المواقع، التي تم رصدها، بخريطة الهوائيات التابعة للوكالة الحكومية (http://emf2.bundesnetzagentur.de/karte/default.aspx) يمكننا تحديد موقعه في كل دقيقة من 260640 دقيقة انقضت طوال 181 يوما ووضعها عبر 'انترفيس' لبرمجة التطبيق عبر خرائط جوجل. وبالتعاون مع فريق الرسوم التخطيطية (الغرافيكس) والتصميم لدينا في صحيفة دي تسايت الإلكترونية، تمكنّا من ابتكار وصلة إنترفيس مناسبة لتصفح البيانات: فمن خلال ضغط زر التشغيل، ستنطلق في رحلة عبر حياة "مالتي سبيتز". هذا هو نوع العمليات التي من الأرجح أن يستخدمها أي صحفي جيد ليصنع ملفاً شخصياً لأي شخص تحت المراقبة.

بعد الإطلاق الناجح للمشروع في ألمانيا، لاحظنا أننا نحصل على عدد كبير جداً من الزيارات من خارج ألمانيا وقررنا أن نصنع نسخة إنجليزية من التطبيق. وبعد أن فزنا بجائزة غريم للصحافة الإلكترونية، تم تكريم المشروع بجائزة مؤسسة الصحافة الإلكترونية (ONA) في سبتمبر 2011، وهي المرة الأولى التي يحصل على هذه الجائزة موقع إلكتروني ألماني للأخبار.

جميع البيانات متاحة في جدول بياني من مستندات جوجل (http://bit.ly/zeitonline-data).

إقرأ الخبر على موقع صحيفة دي تسايت الإلكترونية (http://www.zeit.de/datenschutz/malte-spitz-data-retention).

— ساشا فينور، صحيفة دي تسايت الإلكترونية

أي طراز سيارات؟ معدلات الرسوب في اختبار وزارة النقل

في يناير 2010، حصل موقع بي. بي. سي. على بيانات عن معدلات النجاح والرسوب في الفحص الفني التابع لوزارة النقل لمختلف طرازات السيارات. هذا

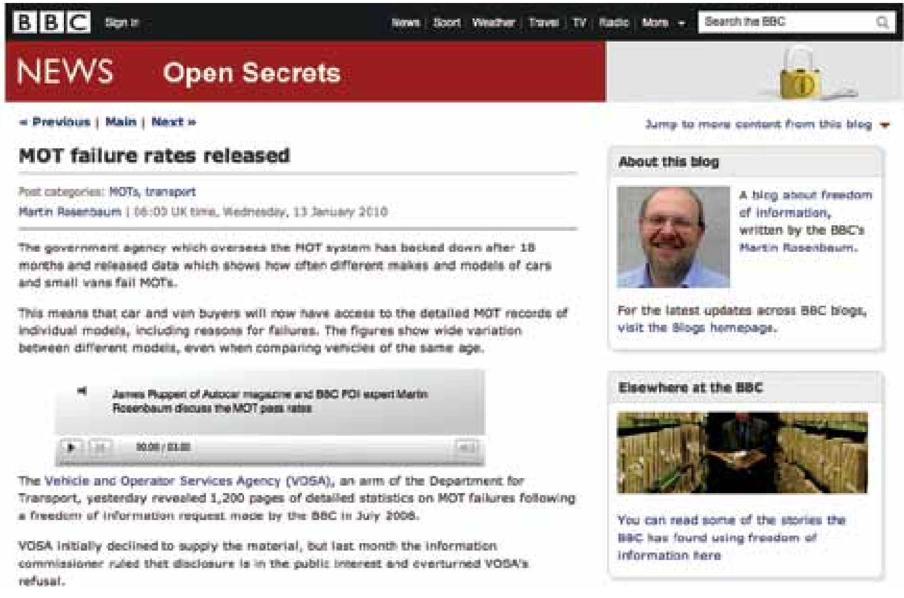
هو الاختبار الذي يقيّم ما إذا كانت السيارة آمنة وملائمة للطريق، وعلى أي سيارة يزيد عمرها عن ثلاث سنوات أن تخضع لاختبار وزارة النقل سنوياً.

حصلنا على البيانات بموجب قانون حرية المعلومات بعد معركة دامت 18 شهراً ضد وكالة خدمات السيارات (VOSA) ووزارة النقل التي تشرف على نظام الاختبار. رفضت الوكالة طلب حرية المعلومات الذي تقدمنا به للحصول على تلك الأرقام بناء على أنها تخرق السرية التجارية. وقد قالت الوكالة أنه قد يقع ضرر تجاري على صناع السيارات بسبب معدلات الرسوب العالية. ولكننا تقدمنا بعد ذلك بالاستئناف إلى مفوض المعلومات الذي حكم بأن نشر المعلومات سيكون للصالح العام. فقامت وكالة VOSA بنشر المعلومات بعد 18 شهراً من تقدمنا بطلب الحصول عليها.

قمنا بتحليل الأرقام مع التركيز على أكثر النماذج رواجاً ومع مقارنة السيارات من نفس العمر، مما أظهر تناقضات واسعة النطاق. فمثلاً، من ضمن السيارات التي بلغ عمرها ثلاث سنوات، رسبت 28٪ من سيارات رينو ميغان في اختبار وزارة النقل في مقابل11٪ فقط من سيارات تويوتا كورولا. أُذيعت الأرقام على التلفزيون والراديو ونُشرت على الإنترنت.

تلقينا البيانات في صيغة وثيقة PDF من 1,200 صفحة، فاضطررنا إلى تحويلها إلى جدول بياني لنتمكن من تحليلها. وبالإضافة إلى نشر نتائجنا، نشرنا هذا الجدول البياني من إكسل (الذي احتوى على ما يزيد عن 14,000 سطر من البيانات) على الموقع الإلكتروني لأخبار BBC إلى جانب الخبر (http://bbc.in/mot-failure-rates)، مما أتاح للجميع الحصول على البيانات في صيغة قابلة للاستخدام.

نتج عن ذلك قيام آخرين باستخدام تلك البيانات لإجراء تحليلاتهم الخاصة والتي لم يسمح لنا الوقت بإجرائها أثناء تعجلنا لنشر الخبر سريعاً أو التي كان من شأنها في بعض الحالات أن تزيد من الضغوط على قدراتنا التقنية في ذلك الوقت. وقد تضمنت تحليلاتهم فحص معدلات رسوب السيارات ذات الأعمار المختلفة، مع مقارنة سجلات الصناع بدلاً من النماذج الفردية وإنشاء قاعدة بيانات قابلة للبحث لإتاحة الفرص للبحث عن نتائج النماذج الفردية. وقد أضفنا روابط لتلك المواقع إلى الخبر على الإنترنت على موقعنا ليستفيد قراؤنا من هذا العمل.


BBC Sign in | News | Sport | Weather | Travel | TV | Radio | More | Search the BBC

NEWS Open Secrets

« Previous | Main | Next »

Jump to more content from this blog

MOT failure rates released

Post categories: MOTs, transport

Martin Rosenbaum | 06:03 UK time, Wednesday, 13 January 2010

The government agency which oversees the MOT system has backed down after 18 months and released data which shows how often different makes and models of cars and small vans fail MOTs.

This means that car and van buyers will now have access to the detailed MOT records of individual models, including reasons for failures. The figures show wide variation between different models, even when comparing vehicles of the same age.

James Ruppert of Autocar magazine and BBC FOI expert Martin Rosenbaum discuss the MOT pass rates

The Vehicle and Operator Services Agency (VOSA), an arm of the Department for Transport, yesterday revealed 1,200 pages of detailed statistics on MOT failures following a freedom of information request made by the BBC in July 2008.

VOSA initially declined to supply the material, but last month the information commissioner ruled that disclosure is in the public interest and overturned VOSA's refusal.

About this blog

A blog about freedom of information, written by the BBC's Martin Rosenbaum.

For the latest updates across BBC blogs, visit the Blogs homepage.

Elsewhere at the BBC

You can read some of the stories the BBC has found using freedom of information here

نموذج 3–20. نشر معدلات رسوب اختبار وزارة النقل (بي. بي. سي.)

أظهر ذلك بعض مميزات نشر البيانات الخام إلى جانب خبر مبني على البيانات. قد تكون هناك استثناءات (إن كنت تنوي مثلاً استخدام البيانات لأخبار متابعة أخرى لاحقاً وترغب في الاحتفاظ بها لنفسك في الوقت الحالي)، ولكن بشكل عام هناك العديد من الفوائد الهامة لنشر البيانات:

- يقتضي عملك أن تكتشف أموراً وتطلع الناس عليها. إذا تحملت عناء الحصول على كل البيانات، فإن نقلها إلى الناس يُعتبر جزءاً من عملك.
- قد يلاحظ آخرون نقاطا هامة لم تلحظها أنت، أو مجرد تفاصيل تهمهم حتى وإن لم تكن هامة بما يكفي لتذكرها في خبرك.
- يمكن للآخرين أن يضيفوا إلى عملك تحليلات للبيانات أعمق وأكثر تفصيلاً، أو أساليب مختلفة لعرض أو توضيح الأرقام مرئياً باستخدام أفكارهم أو مهاراتهم التقنية الخاصة والتي قد تبحث في البيانات بشكل مثمر وبطرق بديلة.
- إنه جزء من دمج المساءلة والشفافية في عملية الصحافة. يمكن للآخرين أن يفهموا أساليبك وأن يتحققوا من عملك إذا أرادوا ذلك.

— مارتن روزنباوم، بي. بي. سي.

دعم الحافلات في الأرجنتين

منذ عام 2002 ظل الدعم المالي لنظام النقل بالحافلات العامة في الأرجنتين ينمو بشكل متزايد، ليحقق رقماً قياسياً جديداً كل عام. ولكن في عام 2011، بعد الفوز بالانتخابات، أعلنت الحكومة الأرجنتينية الجديدة عن تخفيضات في دعم الخدمات العامة ابتداء من شهر ديسمبر من نفس العام. وفي نفس الوقت، قررت الحكومة الوطنية أن تنقل إدارة خطوط الحافلات المحلية وخطوط المترو إلى السلطات في مدينة بوينوس آيريس. ولأنه لم يتم توضيح نقل الدعم إلى تلك السلطة المحلية، وبسبب نقص الأموال المحلية الكافية لضمان سلامة نظام النقل، رفضت حكومة مدينة بوينوس آيريس هذا القرار.

أثناء حدوث ذلك، كنت أنا وزملائي في صحيفة لا ناسيون نجتمع للمرة الأولى لنناقش كيفية بدء عملية صحافة البيانات الخاصة بنا. وقد اقترح محرر القسم المالي أن بيانات الدعم التي نشرتها وزارة النقل (http://www.transporte.gov.ar/) ستمثل تحدياً جيداً للبدء لأن فهمها كان صعبا للغاية بسبب صيغتها ومصطلحاتها.

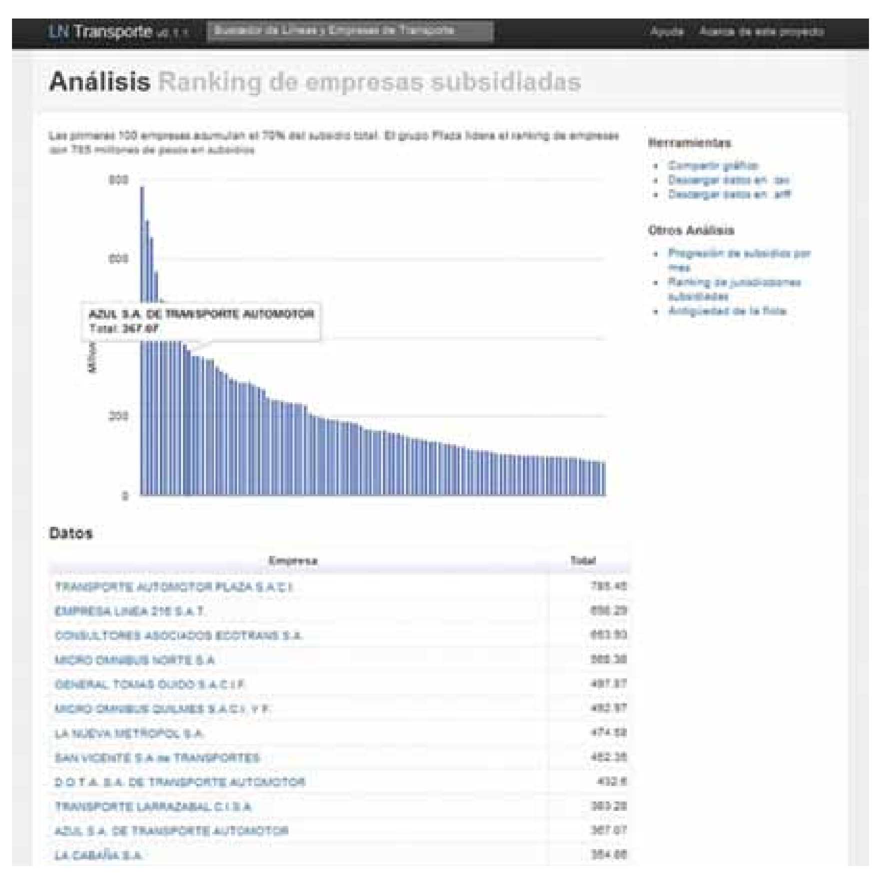
تؤثر الحالة السيئة لنظام النقل العام على حياة أكثر من 5,800,000 راكب يوميا. وغالباً ما تحدث التأخيرات أو الإضرابات أو الأعطال أو حتى الحوادث. فقررنا أن نبحث في نقطة أين يذهب الدعم المخصص لنظام النقل العام في الأرجنتين وأن نتيح تلك البيانات بسهولة لجميع المواطنين الأرجنتينيين من خلال 'مستكشف دعم النقل' وهو مشروع في طور التنفيذ حالياً.

بدأنا بحساب الأموال التي تتلقاها شركات الحافلات شهريا من الحكومة. ولعمل ذلك، نظرنا في البيانات المنشورة على الموقع الإلكتروني لوزارة النقل (http://www.transporte.gov.ar/content/subsidios-sistau/) حيث تم نشر أكثر من 400 وثيقة PDF تحتوي على المدفوعات النقدية الشهرية لأكثر من 1300 شركة منذ عام 2006.

شكلنا فريقا مع أحد كبار المبرمجين لتطوير أداة استخراج بيانات حتى نجعل تحميل وثائق PDF وتحويلها إلى ملفات إكسل وقواعد بيانات تلقائياً. نستخدم الآن مجموعة البيانات الناتجة وبها أكثر من 285,000 سجل لتحقيقاتنا وتوضيحاتنا

المرئية، مطبوعة وعلى الإنترنت. إضافة إلى ذلك، نتيح تلك البيانات في صيغة مقروءة آلياً ليتمكن كل مواطن أرجنتيني من إعادة استخدامها وتعميمها.

كانت الخطوة التالية أن نحدد كم تكلف الصيانة الشهرية لعربات النقل العام الحكومة في المتوسط. ولنكشف عن ذلك، قمنا بزيارة موقع إلكتروني حكومي آخر، وهو موقع اللجنة الوطنية لتنظيم النقل (CNRT) (http://www.cnrt.gov.ar/index2.htm)، أي اللجنة المسؤولة عن تنظيم النقل في الأرجنتين. على هذا الموقع، وجدنا قائمة بشركات الحافلات التي تملك ما مجموعه 9,000 حافلة، فأنشأنا أداة تسوية لتتيح لنا جمع أسماء شركات الحافلات والربط المتبادل بين مجموعتي البيانات بصورة مترافقة تسمح بالمقارنات.



نموذج 3–21. مستكشف دعم النقل (لا ناسيون)

ولاستكمال العمل احتجنا إلى رقم تسجيل كل مركبة. وقد وجدنا على موقع اللجنة قائمة بالمركبات في كل خط للحافلات لكل شركة وبها أرقام لوحات الترخيص الخاصة بها. تتكون أرقام تسجيل المركبات في الأرجنتين من حروف وأرقام تشير إلى ’عُمر‘ المركبة. فرقم سيارتي مثلاً هو IDF234 ويشير حرف ’I‘ إلى مارس–أبريل 2011. قمنا بتسجيل أرقام لوحات ترخيص الحافلات التي تملكها كل من الشركات المذكورة في القائمة لنعرف متوسط أعمار الحافلات في كل شركة، وبالتالي لنتمكن من إظهار مقدار الأموال التي تذهب لكل شركة ومقارنة المبالغ بناء على متوسط أعمار مركباتها.



نموذج 3–22. ترتيب شركات النقل المدعمة (لا ناسيون)

أثناء تلك العملية، تغير بشكل غامض محتوى وثائق PDF التي نشرتها الحكومة والتي تحتوي على البيانات التي نحتاج إليها، على الرغم من أن عناوين URL وأسماء الملفات ظلت كما هي. أما أحد الأمور التي تغيرت فكان افتقاد بعض ملفات PDF لخانات 'المجموع' العمودية، مما جعل مقارنة المجموع خلال فترة التحقيق بأكملها (2002–2011) مستحيلاً.

عرضنا تلك الحالة في ملتقى الصحافيين والمبرمجين في بوسطن حيث تكرم المطوّر مات بيري بابتكار ما نطلق عليه 'جاسوس ملفات PDF'. فاز هذا التطبيق بتصنيف 'الأكثر إثارة للفضول' في هذا الحدث. ويشير تطبيق 'PDF Spy' (http://gristlabs.com/2011/09/24/pdfspy/) إلى صفحة ويب مليئة بملفات PDF ويرى إن كان محتوى الملفات قد تغير. وقد كتب مات بيري "لا تنخدع بـ 'شفافية الحكومة' مرة أخرى".

من الذي عمل على المشروع؟

عمل على هذا التحقيق فريق مكون من سبعة صحافيين ومبرمجين ومصمم تفاعلي مدة 13 شهراً.

كانت المهارات التي احتجنا إليها في هذا المشروع كالآتي:

- صحافيين لديهم علم بكيفية عمل نظام دعم النقل العام وما هي المخاطر ولديهم أيضاً علم بسوق شركات الحافلات.
- مبرمجاً ماهراً في استخراج البيانات من الويب وتحليل البيانات وتسويتها واستخراجها من ملفات PDF إلى جداول إكسل البيانية.
- إحصائي لإجراء تحليل البيانات والحسابات المختلفة.
- مصمم لصنع توضيحات البيانات المرئية التفاعلية.

ما هي الأدوات التي استخدمناها؟

استخدمنا VBasic للتطبيقات وExcel Macros وTableau Public ومنصة Junar للبيانات المفتوحة بالإضافة إلى Ruby on Rails وإنترفيس برمجة التطبيقات في جداول جوجل البيانية وMysql لتطبيق مستكشف الدعم.

كان تأثير المشروع كبيراً، فقد تلقينا آلاف الزيارات وتم نشر التحقيق في الصفحة الرئيسية للنسخة المطبوعة من صحيفة لا ناسيون.

وقد ساعدنا داخلياً نجاح مشروع صحافة البيانات الأول هذا على عرض قضية إنشاء عملية بيانات لتغطية الصحافة الاستقصائية ولتوفير الخدمات للعامة. نتج عن هذا موقع Data.lanacion.com.ar، وهو منصة ننشر فيها البيانات المتعلقة بمختلف المواضيع التي تهم الجمهور في صيغة مقروءة آلياً.

—— أنجيليكا بيرلاتا راموس، صحيفة لا ناسيون (الأرجنتين)



نموذج 3–23. مقارنة عُمر قوافل الحافلات بجحم الأموال التي تتلقاها من الحكومة (لا ناسيون)

صحافيو البيانات المواطنون

ليست غرف الأخبار وحدها التي تستطيع العمل على أخبار مبنية على البيانات، فنفس تلك المهارات التي تفيد صحافيي البيانات يمكنها أيضاً أن تساعد الصحافيين–المواطنين في الحصول على البيانات الخاصة بمنطقتهم وتحويلها إلى أخبار.

كان هذا الدافع الأساسي لمشروع إعلام المواطنين 'أصدقاء جانواريا' 'Friends of Januaria' (http://amigosdejanuaria.wordpress.com/) في البرازيل، والذي حصل على منحة من مشروع 'الأصوات الصاعدة' 'Rising Voices' (http://rising.globalvoicesonline.org/)، وهو مشروع التوعية التابع لـموقع 'الأصوات العالمية' 'Global Voices' (http://globalvoicesonline.org/)، كما تلقى المشروع دعماً إضافياً من منظمة 'المادة 19' 'Article 19' (http://www.article19.org/). ما بين شهري سبتمبر وأكتوبر من عام 2011، تم تدريب مجموعة شباب من سكان بلدة صغيرة في شمال ولاية ميناس جيرايس – وهي من أفقر مناطق البرازيل – على أساسيات أساليب الصحافة ومراقبة الميزانيات. كما تعلموا كيفية تقديم طلبات حرية المعلومات والحصول على المعلومات المتاحة للعامة من قواعد البيانات الرسمية على الإنترنت.

تعرف جانواريا – وهي بلدة يبلغ تعداد سكانها حوالي 65,000 نسمة – أيضاً بإخفاق السياسيين المحليين بها. ففي ثلاث ولايات مدة كل واحدة منها أربع سنوات، كان هناك سبعة رؤساء بلدية متعاقبين. وقد تمت إقالة معظمهم بسبب مخالفاتهم في الإدارات العامة، بما في ذلك تُهم بالفساد.

غالباً ما تفشل البلدات الصغيرة مثل جانواريا في لفت انتباه الإعلام البرازيلي الذي يركز على المدن الكبرى وعواصم الولايات. ولكن هناك فرصة لأن يصبح سكان البلدات الصغيرة شركاء محتملين في مراقبة الإدارة العامة لأنهم أفضل من يعلم التحديات اليومية التي تواجهها مجتمعاتهم المحلية. ومع وجود الإنترنت كشريك هام آخر، يستطيع السكان الآن الحصول بشكل أفضل على معلومات مثل الميزانية وبيانات محلية أخرى.

بعد المشاركة في اثنتي عشرة ورشة، بدأ بعض الصحافيين المواطنين الجُدد

من جانواريا في إظهار كيفية تنفيذ مفهوم الحصول على البيانات المتاحة للجمهور في البلدات الصغيرة. فمثلاً، كتبت سُريّا أموريم - وهي صحافية مواطنة تبلغ من العمر 22 عاما - خبراً عن عدد الأطباء الذين يحصلون على رواتبهم من المدينة وفقاً لبيانات الحكومة الفيدرالية. ولكنها اكتشفت أن الرقم الرسمي لا يوافق العدد الفعلي في البلدة. ولكتابة هذا المقال، حصلت سُريّا على بيانات الصحة المتاحة على الموقع الإلكتروني للنظام الصحي (SUS) (-http: / /bit.ly /tabnet-data sus)، وهو برنامج فيدرالي يوفر المساعدة الطبية المجانية للشعب البرازيلي. ووفقاً لبيانات النظام الصحي، يجب أن يكون في بلدة جانواريا 71 طبيبا في مختلف التخصصات.

لم يتساو عدد الأطباء المذكورين في بيانات النظام الصحي مع ما تعلمه سُريّا عن الأطباء في تلك المنطقة: فالسكان دائماً يشكون نقص الأطباء وقد اضطر بعض المرضى إلى السفر إلى بلدات مجاورة لاستشارة الطبيب. أجرت سُريّا لاحقا مقابلة مع امرأة كانت قد أصيبت في حادث بدراجة نارية ولم تستطع الحصول على مساعدة طبية في مستشفى جانواريا لأنها لم تجد أي طبيب. كما تحدثت سُريّا إلى مسؤول الصحة في البلدة الذي اعترف بوجود أطباء في البلدة أقل من الرقم المنشور في بيانات النظام الصحي.

تثير تلك النتائج الأولية العديد من الأسئلة حول أسباب هذا الاختلاف بين المعلومات الرسمية المنشورة على الإنترنت وبين واقع البلدة. أحد تلك الأسباب هو أن البيانات الفيدرالية قد تكون خطأ، مما يعني أن هناك نقصاً هاماً في معلومات الصحة في البرازيل. هناك احتمال آخر وهو أن جانواريا تنقل معلومات خطأ إلى النظام الصحي. ولكن خبر سُريّا يُعتبر جزءاً هاماً من هذه السلسلة لأنه يسلط الضوء على التناقض وقد يشجع آخرين أيضاً على التدقيق في هذه القضية.

قالت سُريّا "كنت أسكن في الريف، وقد أتممت دراستي الثانوية بصعوبة بالغة. عندما يسألني الناس عما أود فعله في حياتي، كنت دائماً ما أقول لهم أنني أود أن أصبح صحافية. ولكنني تخيلت أن هذا مستحيل تقريباً بسبب العالم الذي أعيش فيه." بعد المشاركة في تدريب ’أصدقاء جانواريا‘، تؤمن سُريّا الآن بأن الحصول على البيانات أداة هامة لتغيير واقع بلدتها. وأضافت "أشعر أنني قادرة



نموذج 3–24. يقوم مشروع إعلام المواطنين 'أصدقاء جانواريا' بتعليم المواطنين المهارات الأساسية لتحويلهم إلى صحافيي بيانات (أصدقاء جانواريا)

على تغيير بلدتي وبلدي والعالم."

هناك صحافي مواطن آخر من هذا المشروع اسمه أليسون مونتيريتون ويبلغ عمره 20 عاماً، وقد استخدم هو أيضاً البيانات لكتابة مقال. كان هذا أثناء الدرس الأول من المشروع، عندما سار الصحافيون المواطنون في أنحاء البلدة باحثين عن مواضيع قد تتحول إلى أخبار. عندها قرر أليسون أن يكتب عن إشارة مرور معطلة تقع في تقاطع شديد الأهمية، وقد ظلت تلك الإشارة معطلة منذ بداية العام. وبعد أن تعلم كيفية البحث عن البيانات على الإنترنت، بحث عن عدد السيارات الموجودة في البلدة وكم الضرائب التي يدفعها مُلاك السيارات. وكتب أليسون:

"يزداد الوضع في جانواريا سوءاً بسبب ارتفاع عدد السيارات في البلدة. ووفقاً للمعهد البرازيلي للجغرافيا والإحصاء (IBGE) (وهو أهم معهد أبحاث إحصائية في البرازيل)، كانت هناك 13,771 مركبة في جانواريا (منها 7,979 دراجة نارية) عام 2010.... يعتقد سكان البلدة أن التأخير في إصلاح إشارة المرور ليس نتيجة

نقص الموارد. ووفقاً لأمانة الموازنة في ولاية ميناس جيرايس، فقد تلقت البلدة 470 ألف ريال برازيلي من ضرائب السيارات عام 2010."

وبالحصول على البيانات، استطاع أليسون إظهار أن جانواريا بها الكثير من السيارات (تقريباً بنسبة سيارة مقابل كل خمسة من السكان) وأن إشارة المرور المعطلة قد تعرّض الكثيرين إلى الخطر. إضافة إلى ذلك، تمكن أليسون من إطلاع جمهوره على حجم الأموال التي تتلقاها البلدة من الضرائب التي يدفعها مُلاك السيارات. وبناء على ذلك، تمكن من التساؤل عما إذا كانت تلك الأموال لا تكفي لإصلاح إشارة مرور لتهيئة ظروف أكثر أمانا للسائقين والمُشاة.

على الرغم من أن هذين الخبرين بقلم سُريّا وأليسون بسيطان للغاية، إلا أنهما يظهران أن الصحافيين المواطنين بإمكانهم استخدام البيانات. لا تحتاج إلى أن تكون في غرفة أخبار كبيرة مع الكثير من المتخصصين لتستخدم البيانات في مقالاتك. بعد اثنتي عشرة ورشة، استطاع كل من سُريّا وأليسون – بدون أي خلفية صحفية لأي منهما – أن يعملا على أخبار مبنية على البيانات ويكتبا مقالات مثيرة للاهتمام عن وضعهم المحلي. إضافة إلى ذلك، تظهر مقالاتهما أن البيانات في حد ذاتها يمكن أن تكون مفيدة حتى وإن كان ذلك على نطاق ضيق. بعبارة أخرى، هناك أيضاً معلومات قيمة في مجموعات البيانات والجداول الصغيرة، وليس فقط في مجموعات البيانات الضخمة.

—— أماندا روسي، أصدقاء جانواريا

اللوحة الكبيرة لنتائج الانتخابات The Big Board

توفر نتائج الانتخابات فرصا رائعة لأي مؤسسة إخبارية لسرد الأخبار مرئياً، ولكن لعدة سنوات كانت تلك فرصة ضائعة بالنسبة لنا. وفي عام 2008، شرعنا مع مكتب الغرافيكس في تغيير ذلك.

أردنا أن نجد طريقة لعرض النتائج التي تسرد قصة ولا تبدو ككتلة من الأرقام في جدول أو خريطة. وهذا بالضبط ما فعلناه في الانتخابات السابقة (http://nyti.ms/senate-1http://nyti.ms/senate-2; http://nyti.ms/senate-3;).

ولا يوجد بالضرورة خلل في كتل الأرقام الكبيرة، أو ما أطلق عليه 'نموذج CNN' للجداول والمزيد من الجداول. فهي ناجحة لأنها تعطي القارئ ما يريد معرفته بالضبط: من الفائز؟

ثمة خطر كبير في العبث بشيء لا يحتاج لإصلاح لأنه ليس معطلاً تماماً، فمن خلال عمل شيء مختلف جذريا والبُعد تماماً عما يتوقعه الناس، من الممكن جعل الأمور أكثر إرباكاً وليس أقل.

في النهاية، توصل شان كارتر من مكتب الغرافيكس إلى الإجابة الصحيحة، وهي ما أطلقنا عليها في نهاية الأمر 'اللوحة الكبيرة' 'The Big Board' (http://nyti.ms/board-elections). وعندما رأيت النماذج للمرة الأولى، كانت لحظة فارقة بحق.

فقد كان محقا تماماً.

نموذج 3–25. اللوحة الكبيرة لنتائج الانتخابات (صحيفة نيويورك تايمز)

ما الذي يجعل منه مثالاً رائعاً على الصحافة المرئية؟ أولاً، تنجذب عين القارئ فوراً إلى الشريط الكبير في الأعلى الذي يوضح أصوات المجمع الانتخابي، وهذا ما قد نطلق عليه في سياق الصحافة اسم (المقدمة). فهي تخبر القارئ بما يريد

معرفته بالضبط وبسرعة وبساطة وبدون أي ضجة بصرية.

الأمر التالي الذي يجذب القارئ تصنيف الولايات في خمسة أعمدة في الأسفل، مرتبة حسب توقع صحيفة نيويورك تايمز لتصويت ولاية معينة لأي من المرشحين. وفي عمود المنتصف ما قد نطلق عليه في سياق الصحافة اسم 'التلخيص'، حيث نشرح سبب فوز أوباما. يجعل العرض التفاعلي هذا واضحا: فاز أوباما بأصوات كل الولايات التي كان من المتوقع أن يفوز بها بالإضافة إلى فوزه بأربع من الخمس ولايات التي كانت فرص الفوز فيها متعادلة مع المرشح الآخر.

بالنسبة لي، هذا الهيكل المكون من خمسة أعمدة يعتبر مثالاً على كيفية اختلاف الصحافة المرئية عن أشكال التصميم الأخرى. من الناحية المثالية، يكون عمل الصحافة المرئية متميزاً إذا كان جميلاً ومفيداً. ولكن عند الاختيار بين الخبر والجماليات، يجب على الصحفي أن يختار جانب الخبر. وعلى الرغم من أن هذا التصميم قد لا يكون الطريقة المفضلة التي يختارها المصمم المتمرس لعرض البيانات، إلا أنه يسرد الخبر بشكل جيد جداً.

وأخيراً، مثل أي عرض تفاعلي جيد على الويب، يدعو هذا العرض القراء إلى المزيد من التعمق. فهناك تفاصيل مثل نسب التصويت في كل ولاية وعدد الأصوات الانتخابية ونسبها، وهذه معروضة عن عمد بقدر أقل من الأهمية حتى لا تنافس النقاط الرئيسية للخبر.

كل هذا يجعل 'اللوحة الكبيرة' عملاً مميزاً في الصحافة المرئية ويرسم خريطة مثالية لأسلوب الهرم المقلوب الجدير بالثقة.

— آرون بيلهوفر، صحيفة نيويورك تايمز

معرفة أسعار المياه عن طريق التكليف (التعهيد) الجماعي

منذ مارس 2011 يتم تجميع معلومات حول سعر مياه الصنبور في أنحاء فرنسا من خلال تجربة في التعهيد أو التكليف الجماعي. وفي أربعة أشهر فقط، قام أكثر من 5,000 شخص ممن ضاقوا ذرعا بتحكم الشركات في سوق المياه بقضاء بعض من وقتهم في البحث عن فواتير المياه الخاصة بهم ومسحها وتحميلها على مشروع 'سعر المياه' (http://www.prixdeleau.fr/). كانت النتيجة تحقيقا غير

مسبوق جمع بين الخبراء والمنظمات غير الحكومية والإعلام التقليدي لتحسين الشفافية بخصوص مشاريع المياه.

تتكون سوق مرافق المياه الفرنسية من أكثر من 10,000 زبون (تقوم المدن بشراء المياه لتوزيعها على الناس) وعدد محدود من شركات المرافق. إن توازن القوى في هذا الاحتكار مائل لصالح الشركات التي تطلب أسعاراً مختلفة أحياناً من البلدات المتجاورة!

تتعامل المنظمة الفرنسية غير الحكومية 'حريات فرنسا' 'France Libertés' مع قضايا المياه حول العالم منذ 25 عاما. تركز المنظمة الآن على تحسين الشفافية في السوق الفرنسية وتمكين المواطنين ورؤساء البلديات الذين يتفاوضون في صفقات مرافق المياه. قررت الحكومة الفرنسية التعامل مع المشكلة منذ عامين من خلال إحصاء على المستوى الوطني. وحتى الآن، تم جمع 3٪ فقط من البيانات. للإسراع من تلك العملية، أرادت المنظمة 'حريات فرنسا' (http://www.france-libertes.org/) إشراك المواطنين بشكل مباشر.

بالتعاون مع فريق موقع OWNI، قمت بتصميم 'إنترفيس' للتعهيد الجماعي حيث يمسح المستخدمون فواتير المياه الخاصة بهم ويكتبون السعر الذي دفعوه مقابل مياه الصنبور في موقع 'سعر المياه' (http://www.prixdeleau.fr/). وفي الأشهر الأربعة الماضية، شارك نحو 8,500 شخص وتم تحميل أكثر من 5,000 فاتورة والتحقق منها.

وعلى الرغم من أن هذا لا يسمح بتقييم مثالي لوضع السوق، إلا أنه أظهر لأصحاب المصلحة مثل الهيئات الوطنية لمراقبة المياه أن هناك اهتماماً حقيقياً على مستوى الجمهور بسعر مياه الصنبور. كانوا في البدء يشككون في الشفافية، ولكنهم غيروا آراءهم أثناء العملية، ثم انضموا تدريجيا إلى نضال منظمة 'حريات فرنسا' ضد غموض الشركات وإهمالها. ما الذي يمكن أن تتعلمه المؤسسات الإعلامية من ذلك؟

الشراكة مع المنظمات غير الحكومية

تحتاج المنظمات غير الحكومية كماً هائلاً من البيانات لتقوم بتصميم وثائق

السياسات العامة. وستكون هذه المنظمات على استعداد أكبر لتمويل عملية جمع البيانات بدلاً من الدفع لمدير تنفيذي لصحيفة ما.

يمكن للمستخدمين أن يوفروا البيانات الخام

تظهر أفضل نتائج التعهيد الجماعي عندما يقوم المستخدمون بمهمة جمع البيانات أو تنقيتها.

أطلب المصدر

فكرنا كثيراً فيما إذا كنا سنطلب من المستخدمين مسح الفاتورة الأصلية، وفكرنا أن هذا سيحول دون اشتراك بعضهم (خاصة أن جمهورنا المستهدف كان أكبر سناً من المتوسط). وعلى الرغم من أن هذا قد أقصى البعض، إلا أنه زاد من مصداقية البيانات.

ضع آلية للتحقق

قمنا بتصميم نظام نقاط وآلية للمراجعة من قبل المشاركين على بعضهم البعض (http://www.prixdeleau.fr/valider) للتحقق من مساهمات المستخدمين. وقد ثبت أن هذا الأمر في غاية التعقيد بالنسبة للمستخدمين الذين كان حافزهم ضعيفاً للدخول إلى الموقع بشكل متكرر. ولكن تم استخدامه من قِبل فريق منظمة 'حريات فرنسا' الذي كان نظام النقاط بمثابة حافز لأعضائه البالغ عددهم حوالي 10 أشخاص.

أبق الأمر مبسطا

بنينا نظاماً آلياً للمراسلات ليتمكن المستخدمون من تقديم طلب حرية المعلومات بشأن أسعار المياه من خلال بضع نقرات فقط. وعلى الرغم من أن هذه الخاصية مبتكرة ومصممة جيداً، إلا أنها لم توفر عائداً كبيراً (فقد تم إرسال 100 طلب فقط).

استهدف جمهورك

أقامت منظمة ’حريات فرنسا‘ شراكة مع مجلة ’60 مليون مستهلك‘ ’60 Millions de Consommateurs‘ **الخاصة بحقوق المستهلكين، والتي تقوم بإشراك مجتمعها بشكل كبير. وقد كانت هذه أفضل شراكة لمثل تلك العملية.**

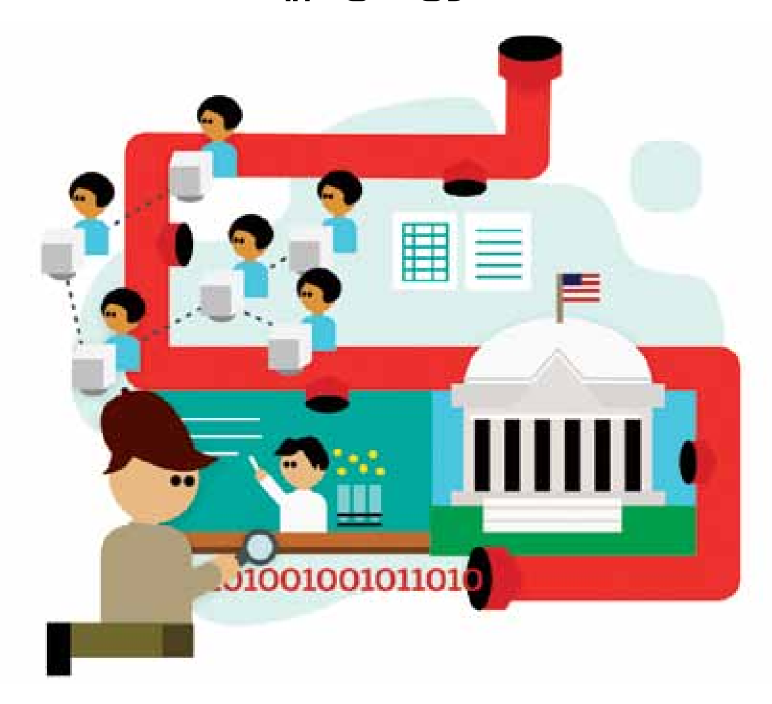
اختر مؤشرات الأداء الأساسية بحرص

جمع المشروع 45,000 زائر في أربعة أشهر فقط، مما يعادل عدد زيارات موقع نيويورك تايمز (/http://www.nytimes.com) في 15 دقيقة. ولكن ما يهم حقا هو أن واحداً من أصل كل خمسة اشترك وأن واحداً من أصل كل عشرة قضى بعض من وقته ليمسح ويحمّل فاتورته.

—— نيكولاس كايسر – بريل، شبكة ++Journalism

الفَصْلُ الرّابع

الحصول على البيانات



هل أنت مستعد للبدء في أول مشروعاتك في صحافة البيانات؟ ما العمل الآن؟ أولاً، تحتاج إلى بعض البيانات. هذا الباب يبحث في أماكن الحصول على البيانات، فستتعلم كيفية الحصول عليها من الإنترنت، وكيفية طلبها باستخدام قوانين حرية المعلومات، واستخدام تقنية 'استخراج البيانات' (screen–scraping) لجمع البيانات من مصادر غير منظمة. كما ستتعلم كيفية استخدام التعهيد الجماعي لتجميع مجموعات البيانات من قرائك. وسننظر فيما تنص عليه القوانين بشأن إعادة نشر مجموعات البيانات، وكيفية استخدام الأدوات القانونية البسيطة حتى تسمح للآخرين باستخدام بياناتك. ويختتم هذا الباب ببعض الحكايات الطريفة وقصص الحرب وبعض ما تعرض له بعض المساهمين في سبيل الحصول على بيانات كانوا يبحثون عنها.

دليلك الميداني في خمس دقائق

هل تبحث عن بيانات موضوع أو مسألة معينة؟ لست متأكدًا إذا كان هناك بيانات أو أين تجدها؟ لا تعلم من أين تبدأ؟ ستتعلم في هذا الباب كيفية البدء في العثور على مصادر البيانات العامة على الويب.

سلاسة البحث

على الرغم من صعوبة العثور على قواعد البيانات على الويب في أغلب الأحوال، إلا أن محركات البحث تفهرس العديد منها، سواء أرغب الناشر في ذلك أم لا. وفيما يلي بعض النصائح للوصول إلى ذلك:

- عند البحث عن البيانات، تأكدمن وضع كل من مصطلحات البحث المتعلقة بمحتوى البيانات التي تبحث عنها وبعض المعلومات عن النمط أو الصيغة أو المصدر الذي تتوقع ورود البيانات به. يتيح جوجل وبعض محركات البحث الأخرى البحث حسب نوع الملف. فمثلاً، يمكنك البحث عن الجداول البيانية (فقط بإضافة نوع الملف إلى بحثك: XLS أو نوع الملف: CSV)، أو البيانات الجغرافية (نوع الملف: shp) أو مقتطفات من قواعد البيانات (نوع الملف: MDB أو SQL أو DB). ويمكنك البحث عن ملفات PDF إذا اضطررت لذلك (نوع الملف: pdf).
- يمكنك أيضاً البحث بجزء من عنوان URL. بالبحث في جوجل عن 'inurl:downloads filetype:xls' سيحاول محرك البحث أن يعثر على كل ملفات إكسل التي توجد بها كلمة 'downloads' في عنوان الويب الخاص بها (إذا وجدت نتيجة واحدة، فإن الأمر عادة ما يتطلب منك أن تنظر في النتائج الأخرى لنفس المجلد على كمبيوتر خادم الويب Web Server). يمكنك أيضًا أن تحدد نتائج بحثك لتتضمن فقط اسم مجال واحد بالبحث مثلاً عن 'site:agency.gov'.
- هناك حيلة رائجة أخرى وهي عدم البحث عن المحتوى مباشرة، بل عن

أماكن تواجد كتل البيانات. فمثلاً، إذا كتبت في بحثك 'site:agency.gov Directory Listing' فسوف تعثر على بعض القوائم التي وضعها خادم الويب (السيرفر) مع الحصول بسهولة على الملفات الخام الأولية، أما إذا بحثت عن 'site:agency.gov Database Download' فسوف تصل إلى القوائم التي تم وضعها عن قصد.

التوجه مباشرة إلى المصدر

أول حيلة أستخدمها للحصول على البيانات التي تملكها جهة عامة هي محاولة التوجه مباشرة إلى صاحب البيانات، وليس إلى مسؤول الشؤون العامة ولا من خلال قانون حرية المعلومات. يمكنني بالطبع صياغة طلب على أساس قانون حرية المعلومات أو طلب للسجلات العامة، ولكن هذا سيبطئ العملية. ومن المرجح أن أحصل على رد مفاده أن البيانات ليست متاحة في الصيغة التي طلبتها، أو – كما حدث في عدة حالات – بأن الهيئة الحكومية تستخدم برنامجًا خاصًا لا يمكنه استخراج البيانات بالصيغة التي طلبتها. ولكن إذا نجحت أولاً في الوصول إلى الشخص المسؤول عن بيانات تلك المنظمة، أستطيع أن أطرح أسئلة حول نوع البيانات التي يحتفظون بها عن الموضوع وكيف يحفظونها، كما أستطيع أن أعرف صيغتها أيضًا. وينبغي أن أتحدث بلغة البيانات حتى أصِل إلى ما يجب أن أعرفه وأنجح في طلب البيانات. ما هي العوائق أمام هذا الأسلوب؟ عادة ما يصعب الوصول إلى هؤلاء الأشخاص. وسيطلب مني المسؤول الإعلامي أن أتواصل معه. وقد وجدت في تلك الحالات أنه من الأفضل أن أحاول إجراء مكالمة جماعية أو الأفضل من ذلك أن أجتمع شخصيًا مع المسؤول الإعلامي ومسؤول البيانات. وأن أعد للاجتماع بصورة يصعب عليهم رفضها. فأقول إنني "لا أريد أن ألقي بالعمل عليهم ولا أن أطلب منهم طلبًا عامًا أو يشكل عبئًا غير ضروري عليهم، لذا فالاجتماع سيساعدني في فهم ما لديهم بالتحديد وما هي أفضل طريقة يمكنني أن أطلب بها ما أحتاجه بالتحديد."

وإذا لم تُجدِ هذه الطريقة نفعًا، فالبديل هو أن أطلب أولاً تخطيط سجلاتهم وقاموس البيانات الخاص بهم، ثم أطلب البيانات نفسها. أحيانًا أسأل أولاً كيف

وبأي نظام يحفظون البيانات. وبهذا يمكنني البحث في وسائل استخراج البيانات قبل أن أكتب طلبي.

وأخيرًا، حققت أفضل نجاحاتي عندما كنت أعمل في صحيفة صغيرة في مونتانا، حيث كنت في حاجة إلى بعض البيانات عن المقاطعة، وقيل لي إنه لا يمكن استخراجها من الكمبيوتر المركزي، فأجريت بحثًا صغيرًا وعرضت أن أذهب إليهم وأساعدهم. فعمِلتُ مع مسؤول البيانات وأعددنا نصًا قصيرًا ووضعنا البيانات على قرص مرن (كان هذا منذ زمن طويل). وبالتالي، حصلت على بياناتي وأصبحت المقاطعة الآن مُجهَّزة لتوفير البيانات لمن يحتاجها. لم يفعلوا ذلك عن قصد، ولكنهم أيضًا كانوا يحتاجون إلى استخراج البيانات أحيانًا ولم يفهموا نظامهم بالكامل، لذا ساعد بعضنا بعضًا.

— شيريل فيليبس، صحيفة سياتل تايمز

تصفح مواقع البيانات وخدماتها

على مدى السنوات القليلة الماضية، ظهرت بوابات بيانات متخصصة ومراكز بيانات ومواقع بيانات أخرى على الويب، وهي أماكن جيدة للاطلاع على أنواع البيانات الموجودة. وفي البداية يمكن تصفح ما يلي:

بوابات البيانات الرسمية

استعداد الحكومة لنشر مجموعة بيانات معينة يختلف من بلد إلى آخر. الكثير من البلدان تطلق بوابات بيانات (مستوحاة من الموقع الأمريكي data.gov والموقع البريطاني data.gov.uk) وذلك من باب ترويج إعادة استخدام المعلومات الحكومية من قبل القطاعات المدنية والتجارية، ويمكن العثور على فهرس عالمي حديث لتلك المواقع على موقع (http://datacatalogs.org/). وهناك موقع آخر مفيد يُدعى 'بيانات حكومات العالم في صحيفة الغارديان' (http://www.guardian.co.uk/world-government-data)، وهو محرك بحثي يشمل العديد من فهارس بيانات الحكومات الدولية.



نموذج 4–1. موقع Datacatalogs.org (مؤسسة المعرفة المفتوحة)

مركز البيانات (/http://thedatahub.org)

هو مورد بيانات يقوم على إسهامات المجتمعات، وتديره مؤسسة المعرفة المفتوحة مما يسهل العثور على مصادر البيانات المتاحة للجميع ومشاركتها وإعادة استخدامها، خاصة بطرق آلية.

موقع ScraperWiki (http://scraperwki.com/)

هو أداة على الإنترنت "لتسهيل عملية استخراج أجزاء مفيدة من البيانات يمكن إعادة استخدامها في تطبيقات أخرى، أو ليتمكن الصحافيون والباحثون من التنقيب فيها." معظم أدوات استخراج البيانات بالموقع وقواعد البيانات عامة ويمكن إعادة استخدامها.

بوابات البيانات لدى البنك الدولي (/http://data.worldbank.org) والأمم المتحدة (/http://data.un.org)

توفر تلك الخدمات مؤشرات عالية المستوى لكل البلدان، وغالبًا ما تعود إلى عدة سنوات ماضية.

مواقع Buzzdata (/http://buzzdata.com) و Infochimps (.http://www infochimps.com/) و DataMarket (/http://datamarket.com)

توجد مواقع ناشئة تهدف إلى بناء مجتمعات محلية في نطاق مشاركة البيانات وإعادة بيعها. ويمكن المشاركة والتعاون على العمل في مجموعات البيانات الخاصة والعامة بموقع Buzzdata (/http://buzzdata.com)، ومتاجر البيانات على موقعي Infochimps (/http://www.infochimps.com) و DataMarket (/http://datamarket.com).

موقع DataCouch (/http://datacouch.com)

حيث يمكنك التحميل والتنقيح والمشاركة والعرض المرئي للبيانات.

موقع Freebase (/http://www.freebase.com)

شركة تابعة لجوجل وهي مثيرة للاهتمام حيث توفر "كياناً بيانيًا (entity graph) للأشخاص والأماكن والأغراض، بناه مجتمع محلي يعشق البيانات المفتوحة."

بيانات البحث

هناك العديد من المجمعات الوطنية والمنظمة لبيانات البحث مثل أرشيف البيانات لدى المملكة المتحدة (/http://www.data–archive.ac.uk). على الرغم من أن هناك الكثير من البيانات المجانية عند الدخول، إلا أن هناك أيضًا الكثير من البيانات التي تتطلب اشتراكًا ولا يمكن إعادة استخدامها أو إعادة توزيعها بدون طلب الإذن أولاً.

الحصول على البيانات من الأرشيف الورقي

بمجرد نشر موقع ويكيليكس لوثائق الجيش الأمريكي في أفغانستان والعراق، قررنا أن نقتبس الفكرة للاحتفال بالذكرى الخمسين لحرب الجزائر عن طريق نشر مذكرات حرب الجزائر. شرعنا في جمع أرشيف الجيش الفرنسي في الجزائر وتحويله إلى الصيغة الرقمية، والبيانات متاحة في أرشيف وزارة الحرب في باريس، ولكن في صيغة ورقية. وقد أرسلنا صحافيين وطُلابًا ليلتقطوا صورًا للوثائق. كما حاولنا مسح الوثائق باستخدام الماسح الضوئي المحمول Canon

P-150 ولكننا لم نفلح لأن معظم أوراق الأرشيف مدعمة بدبابيس.

في النهاية، استطعنا أن نجمع 10,000 صفحة في غضون بضعة أسابيع. وقد استخدمنا معها برنامج التعرف على النصوص (ABBYY FineReader) مما أسفر عن نتائج ضعيفة. إضافة إلى ذلك، منعت الوزارة تعسفيًا الدخول إلى أكثر الصناديق إثارة للاهتمام في الأرشيف. وفوق كل ذلك، تمنع الوزارة إعادة نشر الوثائق التي يمكن تصويرها بحرية في الموقع، لذا قررنا أن الأمر لا يستحق المخاطرة وعلقنا المشروع.

—— نيكولاس كايسر - بريل، شبكة Journalism++

استعن بالمنتديات

ابحث عن الأجوبة السابقة أو اطرح سؤالاً جديدًا في منتديات 'الحصول على البيانات' مثل (http://getthedata.org/) أو 'كورا' (http://www.quora.com/). فالموقع الأول منتدى أسئلة وأجوبة حيث يمكنك طرح الأسئلة المتعلقة ببياناتك، مثل أماكن العثور على بيانات تتعلق بمسألة معينة، وكيفية الاستعلام عن مصادر بيانات معينة أو استرجاعها، وأي الأدوات تستخدم لاستكشاف مجموعة بيانات في صيغة شكل مرئي، وكيفية تنقية البيانات أو وضعها في الصيغة التي يمكنك العمل بها.

استعن بالقوائم البريدية

تذخر القوائم البريدية بالمعرفة التي اكتسبها مجتمع بأكمله بشأن موضوع معين. وبالنسبة لصحافيي البيانات، تعتبر قائمة data-driven-journalism 'الصحافة المبنية على البيانات' (http://bit.ly/ddj-list) وقائمة NICAR-L (http://bit.ly/nicar-subscribe) نقطتين متميزتين للبداية. فالقائمتان حافلتان بصحافيي البيانات وخبراء الصحافة بمساعدة الكمبيوتر (CAR) الذين يعملون على مختلف أنواع المشروعات. ومن المحتمل أن يكون أحدهم قد عمل على خبر مثل الذي تعمل عليه، وقد تكون لديه فكرة عن نقطة البداية أو رابط للبيانات نفسها. يمكنك أيضًا أن تجرب مشروع Wombat (http://project-wombat.

/org) (‘قائمة مناقشات للأسئلة الصعبة’)، والقوائم البريدية العديدة لدى مؤسسة المعرفة المفتوحة (http://lists.okfn.org/mailman/listinfo)، وقوائم theInfo البريدية (http://theinfo.org/)، أو البحث عن قوائم بريدية خاصة بموضوعك أو مجال اهتمامك.

انضم إلى ملتقى الصحافيين والمبرمجين Hacks/Hackers

ملتقى الصحافيين والمبرمجين هو منظمة صحفية شعبية دولية آخذة في التوسع ولديها عشرات التجمعات وآلاف الأعضاء في أربع قارات. تهدف المنظمة إلى إنشاء شبكة من الصحافيين (hacks) والمبرمجين التقنيين (hackers) ليقوموا بإعادة التفكير في مستقبل الأخبار والمعلومات. ومن خلال هذه الشبكة الواسعة، يوجد احتمال كبير بأن تجد من يعرف مكان البحث عما تسعى إليه.

استعن بالخبراء

عادة ما يعرف الأساتذة والموظفون الحكوميون وأهل الصناعة أماكن البحث. اتصل بهم وراسلهم عبر البريد الإلكتروني، وتحدث معهم في الفعاليات، واذهب إلى مكاتبهم واسألهم بلطف: "أنا أعدُّ تقريرًا حول موضوع كذا، فأين يمكنني أن أجد البيانات؟ هل تعرف من يدلني إلى ذلك؟"

تعرّف على تقنية المعلومات المستخدمة لدى الحكومة

عند محاولة الوصول إلى البيانات، في كثير من الأحيان، من المفيد أن تفهم السياق التقني والإداري حيث تحفظ الحكومات معلوماتها. فبغض النظر سواء أكانت قاعدة بيانات CORDIS أم COINS أم THOMAS– قواعد البيانات ذات الأسماء المختصرة بكتابة الأحرف الأولية للعديد من الكلمات، غالبا ما تكون أكثر إفادة بمجرد فهم القليل عن الغرض المقصود منها.

اعثر على المخططات أو الهياكل التنظيمية للحكومة وابحث عن الأقسام أو الوحدات ذات الخدمات الشاملة (مثل إعداد التقارير أو الخدمات التكنولوجية)، ثم استكشف مواقعها الإلكترونية. يتم حفظ الكثير من البيانات في عدة أقسام، وبينما يتعامل بعض الأقسام مع هذه البيانات على أنها الدُّرّ المصون، ستجد أقسامًا

أخرى توفرها لك بدون مقابل.

انتبه للرسوم البيانية الديناميكية على المواقع الإلكترونية الحكومية. فغالبًا ما تدعمها مصادر بيانات منظمة أو وصلات تداخل لبرمجة التطبيقات API التي يمكن استخدامها بشكل مستقل (مثل التطبيقات الصغيرة لتتبع رحلات الطيران أو تطبيقات جافا للنشرة الجوية).

البحث في سجلات الهاتف

منذ بضعة أشهر، أردت تحليل سجلات هاتف حاكم ولاية تكساس ريك بيري (الذي كان مرشحًا رئاسيًا في ذلك الوقت). كان ذلك نتيجة لطلب سجلات عامة طال انتظارها. جاءت البيانات في أكثر من 120 صفحة من الوثائق من نوعية الفاكس. وقد كان ذلك جهدًا تطلب إدخال البيانات وتنقيتها، ثم تلت ذلك وصلة برمجة للتطبيقات من موقع دليل أرقام الهاتف WhitePages.com للبحث العكسي في أرقام الهواتف.

وبعد أن قمنا بدمج الأسماء مع بيانات الولاية وبيانات الانتخابات الفدرالية، وجدنا أن بيري كان قد تواصل مع المتبرعين للحملات الانتخابية واللجان السياسية الكبرى من هواتف العمل، وهو السلوك المستهجن الذي أثار الأسئلة حول علاقته بإحدى اللجان السياسية الكبرى التي تعمل لصالحه.

—— جاك غيلام، وكالة أسوشيتد برس

ابحث مرة أخرى

عندما تعرف المزيد عما تبحث عنه، ابحث مرة أخرى باستخدام الجمل ومجموعات الكلمات بعيدة الاحتمال التي وجدتها بعد آخر مرة قمت فيها بالبحث. قد يحالفك المزيد من الحظ مع محركات البحث.

اكتب طلباً لحرية المعلومات

إذا كنت تعتقد أن هناك هيئة حكومية معينة لديها البيانات التي تحتاج إليها، قد يكون تقديم طلب حرية المعلومات أفضل أداة لديك. انظر في الباب التالي

للمزيد من المعلومات حول تقديم الطلب.

—— بريان بوير (صحيفة شيكاغو تريبيون)، جون كيف (إذاعة WNYC)، فريدريك ليندنبرغ (مؤسسة المعرفة المفتوحة)، جاين بارك (منظمة المشاع الأبداعي)، كريس وو (ملتقى الصحافيين والمبرمجين / Hacks Hackers)

عندما يخفق القانون

بعدما قرأت مقالاً أكاديميًا (http://bit.ly/hygiene–inspections) يوضح أن نشر نتائج التفتيش الصحي للمطاعم قلّل من معدل انتشار الأمراض المتعلقة بالطعام في لوس أنجلوس، طلبت من الخدمات الصحية في باريس لائحة التفتيش. وبعد إجراءات قانون حرية المعلومات الفرنسي، انتظرت 30 يوماً لرفضهم الرد، ثم ذهبت إلى 'لجنة الحصول على البيانات العامة' (CADA بالفرنسية)، والتي تَفصِل في شرعية طلبات قانون حرية المعلومات. أيدت اللجنة طلبي وأمرت الإدارة بنشر البيانات. فطلبت اللجنة شهرين إضافيين ووافقت اللجنة على ذلك. وبعد انقضاء شهرين، لم تفعل الإدارة شيئًا.

حاولت إقناع بعض الأسماء الكبيرة (والجيوب الكبيرة) من المدافعين عن البيانات المفتوحة بالذهاب إلى المحكمة (وهو ما يكلف 5000 يورو ويعتبر فوزًا مضموناً مع دعم اللجنة)، ولكنهم كانوا يخشون المخاطرة بعلاقتهم مع برامج البيانات المفتوحة الرسمية. وهذا أحد الأمثلة العديدة حيث تتجاهل الإدارة الفرنسية بكل بساطة القانون، ولا تفعل المبادرات الرسمية شيئًا لدعم الطلبات الشعبية للبيانات.

—— نيكولاس كايسر – بريل، شبكة Journalism++

حقك في الحصول على البيانات

قبل أن تقدم طلب حرية المعلومات، عليك أن تبحث إذا ما كانت البيانات متاحة للتداول بالفعل، أو إذا سبق وأن طلبها آخرون. جاء في الباب السابق بعض الاقتراحات بخصوص أماكن البحث. فإذا بحثت ولم تتمكن من الحصول على

البيانات التي تحتاج إليها، فقد تحتاج أن تقدم طلبًا رسميًا. وفيما يلي بعض النصائح التي تساعد في جعل طلبك أكثر فعالية:

خطط مقدما لتوفير الوقت

فكِّر في تقديم طلب رسمي كلما شَرَعتَ في البحث عن معلومات، فمن الأفضل ألا تنتظر حتى نفاد كل الاحتمالات الأخرى. إذا قدمت الطلب في بداية بحثك مع القيام بالتحقيقات الأخرى على التوازي فإنك سوف توفر الوقت. كن مستعدًا للتأخر: فأحيانًا تستغرق الجهات الحكومية وقتًا طويلًا للنظر في الطلبات، لذا من الأفضل أن تتوقع التأخر.

راجع القوانين المتعلقة بالرسوم

ينبغي مراجعة القوانين المتعلقة بالرسوم سواء أكانت مقابل تقديم الطلبات أو تلقّي المعلومات قبل تقديمك للطلب، وبذلك ستعرف حقوقك إذا فاجأك مسؤول حكومي وطلب منك المال. كما يمكنك أيضاً أن تطلب الوثائق الإلكترونية لتفادى النسخ وزيادة النفقات، وعليه اذكر في طلبك أنك تفضّل الحصول على المعلومات في صيغة إلكترونية. وهكذا تتجنب دفع الرسوم، إلا إذا كانت المعلومات غير متاحة إلكترونيًا، على الرغم من أنه يمكن في هذه الأيام مسح الوثائق التي لم يتم تحويلها إلى الصيغة الرقمية بعد، ثم إرفاقها في رسالة عبر البريد الإلكتروني.

اعرف حقوقك

اعرف حقوقك قبل أن تبدأ حتى تحدد موقفك وما يتوجب على الجهات الحكومية فعله أو عدم فعله. فمثلاً، تعطي معظم قوانين حرية المعلومات مُهلة للجهات الحكومية حتى ترد عليك. وفي معظم القوانين على مستوى العالم تتراوح هذه المُهلة من بضعة أيام إلى شهر واحد، فاحرص على معرفة مدة المُهلة قبل أن تبدأ، وقُم بتدوين تاريخ تقديم طلبك.

الحكومة غير ملزمة بمعالجة البيانات من أجلك، ولكنها ملزمة بإعطائك كل ما لديها من بيانات، وإذا كانت لديهم تلك البيانات من أجل الاضطلاع باختصاصاتهم القانونية، فعليهم بالتأكيد أن يوفروا لك هذه البيانات.

بيّن أنك تعرف حقوقك

عادة لا يتطلب القانون أن تذكر قانون حق الحصول على المعلومات أو قانون حرية المعلومات، ولكن يُوصى بذلك لأنه يُظهر أنك على علم بحقوقك القانونية ومن المحتمل أن يشجع ذلك على اتباع الإجراءات الصحيحة في التعامل مع الطلبات وفقًا للقانون. ونلاحظ أنه عند تقديم الطلبات للاتحاد الأوروبي، يجب أن تذكر أنه طلب حق الحصول على الوثائق، ومن الأفضل أن تذكر على وجه الخصوص البند 2001/1049.

اجعله بسيطًا

في جميع البلدان، من الأفضل أن تبدأ بطلب بسيط للمعلومات ثم تضيف المزيد من الأسئلة بمجرد حصولك على المعلومات المبدئية. وهكذا تتجنب مماطلات المؤسسة الحكومية بطلب التمديد في الوقت لكون 'الطلب معقدًا'.

اجعله مُركّزًا

إذا كان طلب المعلومات موجهًا إلى أحد الأقسام في جهة حكومية، فمن المحتمل أن يستغرق الرد وقتًا أقصر مما لو كان الطلب يقتضي البحث في جميع الأقسام. فالطلب الذي يحتم مراجعة الجهة لأطراف أخرى ذات صلة بالمعلومات (مثل شركة خاصة توفر المعلومات أو حكومة أخرى تتأثر بهذه المعلومات) قد يستغرق وقتاً أطول، فكُن مثابرًا.

فكر فيما يوجد بداخل خزانة الملفات

حاول أن تعرف طبيعة البيانات التي يتم جمعها. فمثلاً، إذا حصلت على أحد النماذج الذي يعبئه شرطي المرور حال وقوع حادثة، فستعرف طبيعة المعلومات التي تسجل في مثل هذه الحوادث.

كُن محدداً

قبل أن تقدم طلبك، فكر في الآتي: هل الطلب مُبهم بأي شكل من الأشكال؟ وهذا سؤال هام خاصة إن كنت تنوي عقد مقارنة بين البيانات الصادرة عن مختلف

الجهات الحكومية. فمثلاً، إذا طلبت أرقام السنوات الثلاث الماضية، سترسل لك بعض الجهات المعلومات للسنوات التقويمية الثلاث الماضية، وسترسل لك جهات أخرى السنوات المالية الثلاث الماضية، مما سيمنعك من المقارنة المباشرة بينهما. وإذا قررت أن تخبئ طلبك الحقيقي داخل طلب أكثر شمولاً، فاحرص أن يكون طلبك عامًا بما يكفي لتحصيل المعلومات التي تريدها شريطة ألا يكون عامًا للغاية لدرجة تجعله مُبهمًا أو تُعيق الرد عليه. فغالبًا ما تتلقى الطلبات المحددة والواضحة ردودًا أسرع وأفضل.

قدم عدة طلبات

إذا لم تكن متأكداً أين تقدم طلبك، لا يوجد ما يمنعك من تقديمه إلى جهتين أو ثلاث جهات (أو أكثر) في نفس الوقت. في بعض الحالات، ستعطيك الجهات المتنوعة ردوداً مختلفة، ولكن قد يفيدك ذلك في إعطائك الصورة الكاملة للمعلومات المتاحة حول موضوع تحقيقك.

قدم طلبات دولية

تزداد كل يوم فرص تقديم الطلبات إلكترونيًا، لذا لم يعد محل الإقامة عائقًا. وإليك طريقة أخرى إن لم تكن تسكن في البلد التي تريد أن تقدم الطلب فيها، يمكنك أحياناً أن ترسل الطلب إلى السفارة التي ستنقله بدورها إلى الجهة الحكومية المختصة. ستحتاج إلى أن تسأل السفارة أولاً إن كانت على استعداد للقيام بذلك، فأحيانًا لا يكون موظفو السفارة مدربين على التعامل مع إجراءات قانون حق الحصول على المعلومات، وفي هذه الحالة من الأفضل أن تقدم الطلب مباشرة إلى الجهة الحكومية المختصة.

قدم طلبًا تجريبيًا

إذا كنت تخطط لإرسال طلب إلى عدة جهات حكومية، ابدأ بإرسال مسودة أولية للطلب إلى بعض الجهات من باب التجربة. وبالتالي سيتضح لك هل تستخدم المصطلحات الصحيحة للحصول على المادة التي تريدها وهل من الممكن الرد على سؤالك، وعليه سيمكنك تنقيح الطلب– إن اقتضى الأمر ذلك– قبل أن ترسله للجميع.

استبق وقوع الاستثناءات

إن كنت تظن أن طلبك سيخضع لبعض الاستثناءات، فينبغي عند إعداد الأسئلة أن تفصل بين الأسئلة التي تتعلق بالمعلومات التي قد تكون حساسة وبين المعلومات الأخرى التي من المنطقي ألا تنطبق عليها أي استثناءات. ثم قسم أسئلتك إلى جزأين وقدم طلبين منفصلين.

اطلب الحصول على الملفات

إذا كنت تسكن قريباً من مكان حفظ المعلومات (أي في العاصمة التي تحفظ فيها الوثائق مثلاً)، يمكنك أيضًا أن تطلب معاينة الوثائق الأصلية. قد يكون ذلك مفيدًا عند البحث عن معلومات محفوظة في العديد من الوثائق التي تودّ إلقاء نظرة عليها جميعًا. ينبغي أن تكون تلك المعاينة بالمجان وأن يتم ترتيبها في وقت معقول ومناسب لك.

احتفظ بسجل!

قدم طلبًا كتابيًا واحتفظ بسجل له أو نسخة منه حتى تتمكن مستقبلاً من إثبات تقديمك للطلب، إذا ما أردت أن تتقدم بشكوى من عدم الرد عليك، وبالتالي يكون ذلك دليلاً على تقديم الطلب إذا كنت تنوي نشر خبر عن ذلك.

أخبر الجمهور عن طلبك

سرّع الحصول على الردود بالإعلان على الملأ عن تقديمك لهذا الطلب، إذا كتبت أو أذعت خبرًا عن تقديم الطلب، فقد يمثل ذلك ضغطًا على المؤسسة الحكومية للنظر في الطلب والرد عليه. وعليه يمكنك إحاطة الجمهور بالمعلومات المحدثة حينما تحصل على الرد، أما إذا انتهت المُهلة دون تلَّقي أي رد فيمكنك أن تصنع من تلك الحادثة خبرًا. وهناك فائدة إضافية لذلك وهي توعية الناس بحقهم في الحصول على المعلومات وكيفية ممارسة ذلك الحق.

نصيحة: توجد أيضاً عدة خدمات ممتازة يمكنك استخدامها لإتاحة طلبك وأي ردود عليه للناس على مواقع الويب، مثل موقع 'ماذا يعلمون؟ 'What

Do They Know? (http://www.whatdotheyknow.com/) للجهات الحكومية البريطانية، وموقع ‘اسأل الدولة’ Frag den Staat (https://fragdenstaat.de/) للجهات الحكومية الألمانية، وموقع ‘اسأل الاتحاد الأوروبي’ Ask the EU (http://www.asktheeu.org/) لمؤسسات الاتحاد الأوروبي. ويساعد مشروع Alaveteli (http://www.alaveteli.org/) على تقديم خدمات مشابهة إلى عشرات البلدان حول العالم.



نموذج 4–2. موقع ‘ماذا يعلمون؟’ (**My Society**)

قُم بإشراك الزملاء

إذا كان زملاؤك يشككون في قيمة طلبات الحصول على المعلومات، فإن من أفضل الطرق لإقناعهم كتابة خبر مبني على المعلومات التي حصلت عليها باستخدام قانون الحصول على المعلومات. كما يُوصَى بأن تذكر في نهاية المقالة أو العمل الإذاعي أنك استخدمت القانون، وهذا لتعزيز قيمته ونشر الوعي العام عن هذا الحق.

اطلب البيانات الخام

إذا أردت أن تحلل البيانات أو تستكشفها أو تستخدمها بواسطة الكمبيوتر، فعليك أن تطلب بكل وضوح توفير البيانات في صيغة إلكترونية ومقروءة آليًا. ويمكنك توضيح ذلك بأن تحدد مثلاً أنك تطلب معلومات الموازنة في صيغة 'مناسبة للتحليل بواسطة برنامج محاسبة.' ويمكنك كذلك أن تطلب بوضوح توفير المعلومات في شكل مصنّف أو مُفصّل. تستطيع قراءة المزيد عن هذا الموضوع في هذا التقرير (http://bit.ly/access-report).

السؤال عن المنظمات المستثناة من الالتزام بقوانين حرية المعلومات

قد تود أن تتعرف على المنظمات غير الحكومية أو الشركات الخاصة أو المؤسسات الدينية أو أي مؤسسات أخرى غير ملزمة بنشر الوثائق بموجب قوانين حرية المعلومات. ولكن من الممكن العثور على معلومات بشأن هذه المؤسسات من خلال سؤال الهيئات العامة التي تخضع لقوانين حرية المعلومات. فيمكنك، على سبيل المثال، أن تسأل إحدى الجهات أو الوزارات الحكومية عن التمويل أو التعامل مع شركة خاصة معينة أو منظمة غير حكومية وأن تطلب منهم الوثائق المساندة. وإذا احتجت إلى المزيد من المساعدة في تقديم طلب حرية المعلومات الخاص بك، بإمكانك أن تتصفح 'عدة أدوات التسريبات القانونية للصحافيين' (http://www.legalleaks.info/toolkit.html).

— هيلين داربيشاير (Access Info Europe)، جورجي باديسكي (عضو زمالة Knight Journalism في جامعة ستانفورد)، مارتن روزنبوم (بي. بي. سي.)، فابريتزيو سكروليني (كلية لندن للاقتصاد والعلوم السياسية).

استخدام قانون حرية المعلومات لفهم الإنفاق العام

لقد استخدمت قانون حرية المعلومات ببضع طرائق مختلفة للمساعدة في تغطية قاعدة بيانات COINS، وهي أكبر قاعدة بيانات للإنفاق والموازنة والمعلومات المالية في حكومة المملكة المتحدة. في بداية عام 2010، قال

جورج أوزبورن أنه إذا أصبح وزيرًا للمالية فإنه سينشر قاعدة بيانات COINS لتقديم قدر أكبر من الشفافية في وزارة المالية. في ذلك الوقت، بدا أن التحقيق في البيانات وفي هيكل قاعدة بيانات COINS فكرة جيدة، لذا أرسلت بعض طلبات حرية المعلومات: منها واحد لمخطط قاعدة البيانات (http://bit.ly/wdtk-coins-1)، وواحد للإرشاد الذي يتلقاه العاملون في وزارة المالية عندما يتعاملون مع قاعدة بيانات COINS (http://bit.ly/wdtk-coins-2)، وآخر للعقد الموقع بين الوزارة ومزود قاعدة البيانات (http://bit.ly/wdtk-coins-3)، مما أدى إلى نشر بيانات مفيدة. وقد طلبت أيضًا الحصول على الرمز الكامل للإنفاق في قاعدة البيانات، والذي تم نشره أيضًا (http://bit.ly/wdtk-coins-4). ساعد كل ذلك على فهم قاعدة بيانات COINS عندما أصبح جورج أوزبورن وزيرًا للمالية في مايو 2010 ونشر قاعدة البيانات في يونيو 2010. وقد تم استخدام بيانات COINS في عدد من المواقع الإلكترونية التي تشجع الناس على التحقيق في البيانات، ومنها موقع OpenSpending.org (http://openspending.org/) ومستكشف بيانات COINS في صحيفة الغارديان (http://coins.guardian.co.uk/coins-explorer/search).

بعد المزيد من التحقيق تبين أن هناك جزءًا كبيرًا من قاعدة البيانات مفقود: الحسابات الحكومية الكاملة (WGA) وهي عبارة عن 1,500 مجموعة من الحسابات للجهات التي تمولها الحكومة. استخدمت قانون حرية المعلومات لطلب بيانات تلك الحسابات لعام 2008/ 2009 (http://bit.ly/wdtk-coins-5) ولكن دون جدوى. طلبت أيضاً التقرير من مكتب التدقيق في الحسابات الحكومية الكاملة (WGA)، وكنت على أمل أن يشرح لي أسباب عدم ملاءمتها للنشر. وقد تم رفض هذا الطلب كذلك (http://bit.ly/wdtk-coins-6).

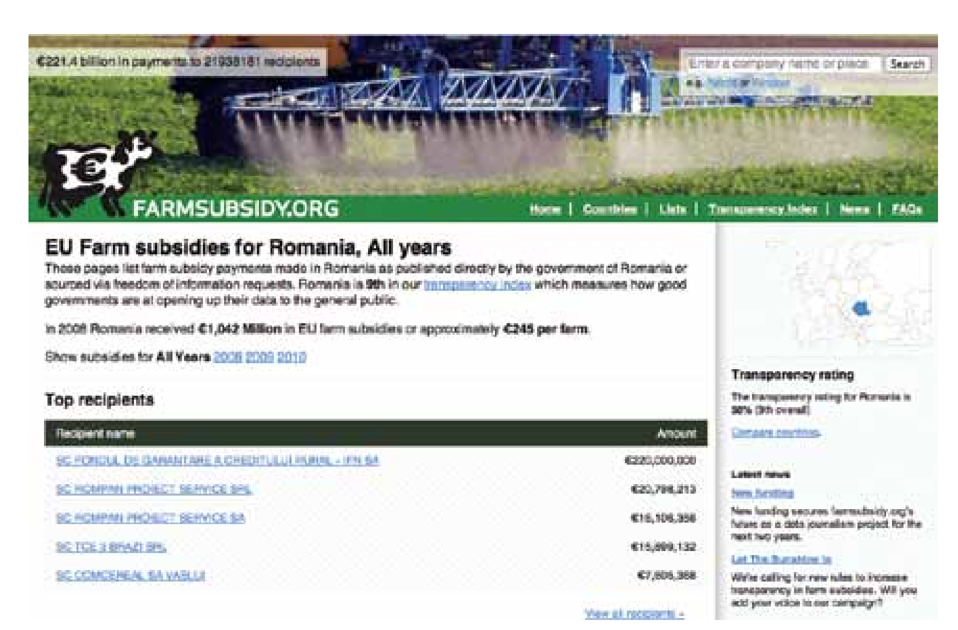
في ديسمبر 2011، نُشرت الحسابات الحكومية الكاملة في قاعدة بيانات COINS. ولكنني أردت أن أتأكد من وجود الإرشاد الكافي لوضع المجموعة الكاملة للحسابات لكل من الألف وخمسمئة جهة، بما في ذلك ممارسات الحسابات الحكومية الكاملة. ويقودنا هذا إلى الطريقة الثانية التي استخدمت

بها قانون حرية المعلومات: وهذا لضمان أن نشر البيانات تحت بند الشفافية في المملكة المتحدة سيكون مشروحاً بشكل جيد ويحتوي على ما يلزم. وقد قدمت طلب حرية المعلومات للحصول على المجموعة الكاملة لحسابات كل الجهات الحكومية المذكورة في الحسابات الحكومية الكاملة (http://bit.ly/wdtk-coins-7).

—— ليزا إيفانز، صحيفة الغارديان

احصل على البيانات باستخدام قانون حرية المعلومات 'Wobbing'

يعد استخدام قانون حرية المعلومات – الذي يدعى أحيانًا 'Wobbing' – أداة ممتازة. ولكنه يتطلب أسلوبًا معينًا وغالبًا ما يتطلب المثابرة كذلك. وفيما يلي ثلاثة أمثلة توضح مميزات وتحديات استخدام قانون حرية المعلومات مستمدة من عملي كصحفي استقصائي.



نموذج 4–3. الموقع الإلكتروني لدعم المَزارع **farmsubsidy.org**

دراسة حالة رقم (1): دعم المَزارِع

يدفع الاتحاد الأوروبي 60 مليون يورو سنويًا للمزارعين ولقطاع الزراعة. ويجري العمل بتقديم هذا الدعم منذ أواخر الخمسينيات وتقول الرواية السياسية إن هذا الدعم لمساعدة المزارعين الأكثر فقرًا. ولكن أول إنجاز لقانون حرية المعلومات في الدنمارك عام 2004 أشار إلى أن تلك مجرد رواية مزعومة. فقد كان المزارعون الصغار يكافحون ويشْكُون في السر والعلن من قلة الدعم. حيث تذهب معظم الأموال إلى بعض كبار مُلاك الأراضي وإلى الصناعات الزراعية. لذا بالطبع أردت أن أعرف: هل هناك نمط سائد في أنحاء أوروبا؟

في صيف عام 2004، طلبت البيانات من المفوضية الأوروبية. حيث تتلقى المفوضية البيانات من الدول الأعضاء في شهر فبراير من كل عام، وتوضح البيانات من يطلب الحصول على تمويل الاتحاد الأوروبي، وكم يتلقى المستفيدون، وسواء أكانوا يحصلون عليه لزراعة أراضيهم وتنمية المنطقة أم لتصدير الحليب المجفف. وفي ذلك الوقت، تلقت المفوضية الأرقام في صيغة ملفات CSV على أقراص مدمجة. وكانت البيانات ضخمة ولكن مبدئيا، كان التعامل معها سهلاً، فكل ما كان عليك فعله هو استخراج البيانات.

في عام 2004 رفضت المفوضية نشر البيانات، وكانت حجتها الرئيسية أن البيانات قد تم تحميلها في قاعدة بيانات ويتطلب استخراجها الكثير من العمل، وهي الحجة التي أطلق عليها أمين المظالم الأوروبي 'سوء إدارة'. بإمكانك العثور على جميع وثائق تلك الحالة من موقع wobbing.eu (http://bit.ly/eu-wobbing). ولكن في عام 2004، لم يكن لدينا الوقت الكافي لنطلب البيانات بشكل قانوني، رغم أننا كنا نريد الحصول عليها.

لذا شكلنا فريقاً مع شركاء من مختلف أنحاء أوروبا للحصول على بيانات كل دولة على حدة. وقد حصل زملاء بريطانيون وسويديون وهولنديون على البيانات في عام 2005. وتمت إتاحة البيانات في فنلندا وبولندا والبرتغال ومناطق في إسبانيا وسلوفينيا وبلدان أخرى أيضًا، وحتى في ألمانيا التي يصعب فيها الحصول على المعلومات، حققت إنجازاً وتلقيت بعض البيانات في مقاطعة شمال الراين-وستفاليا في عام 2007. حيث اضطررت إلى الذهاب إلى المحكمة للحصول

على البيانات، ولكن نتج عن ذلك بضعة مقالات جيدة في مجلة شتيرن الإخبارية المطبوعة والإلكترونية (http://bit.ly/stern-wobbing).

هل كان من الصدفة أن تكون الدنمارك والمملكة المتحدة أول من يتيح البيانات؟ ليس بالضرورة. فبالنظر إلى الصورة السياسية الكاملة، نجد أن دعم المزارع في ذلك الوقت كان لا بد له أن يظهر في سياق مفاوضات منظمة التجارة العالمية حيث كان الدعم يخضع للضغط. ولأن الدنمارك والمملكة المتحدة من بين أكثر البلدان ليبرالية في أوروبا، فقد كانت هناك رياح سياسية تهب في اتجاه الشفافية في تلك البلدان.

لم تنته القصة عند هذا الحد، وللحصول على المزيد من الحلقات والبيانات، تصفح موقع (http://farmsubsidy.org/).

الدرس المستفاد: ابحث عن طرق الحصول على البيانات باستخدام قانون حرية المعلومات، فلدينا منه تنوع رائع في أوروبا، وقد تتنوع المصالح السياسية للعديد من البلدان في أوقات مختلفة، فينبغي استخدام ذلك لصالحك.

اعرف حقوقك

عندما تنشر البيانات، هل عليك أن تقلق إزاء حق النشر والحقوق الأخرى المتعلقة بهذه البيانات؟ ينبغي عليك دائمًا أن تناقش هذا الأمر مع مستشارك القانوني، ولكن كقاعدة عامة: إذا نشرتها الحكومة، لا يجب عليك أن تطلب السماح أو الإذن، وإذا نشرتها منظمة لا تتربح من البيانات، لا تقلق كثيرا، أما إذا نشرتها منظمة تتربح بالفعل من البيانات، فعليك قطعًا أن تطلب الإذن.

— سايمون روجرز، صحيفة الغارديان

دراسة حالة رقم (2): الآثار الجانبية

نحن كلنا فئران تجارب فيما يتعلق بتناول الأدوية. فللعقاقير آثار جانبية، وجميعنا يعلم هذا ولكننا نوازن بين الفوائد والمخاطر المحتملة، ثم نتخذ القرار. وللأسف غالباً ما يكون هذا القرار غير مدروس.

عندما يتناول المراهقون أقراصًا لمعالجة حب الشباب، فإنهم يأملون في

الحصول على بشرة ناعمة وليس مزاجًا سيئًا. ولكن هذا ما نتج عن أحد العقاقير، حيث أصيب الشباب بالاكتئاب وتملكتهم نزعة انتحارية بعد تناول العقار. ولم يكن خطر هذا التأثير الجانبي بالذات جائزًا قانونيًا، مما شكّل بالطبع خبرًا للصحافيين.

توجد بيانات حول الآثار الجانبية. ويجب على المُصنّعين أن يقدموا المعلومات بصورة منتظمة إلى السلطات الصحية بشأن ما لاحظوه من آثار جانبية. وتحتفظ السلطات الوطنية أو الأوروبية بالبيانات بمجرد السماح بتسويق العقار.

ومرة أخرى، تحققت الانفراجة الأولية على المستوى الوطني في الدنمارك، ثم أثناء إجراء بحث عابر للحدود من قبل فريق من الدنمارك وهولندا وبلجيكا، أتاحت هولندا أيضًا بياناتها. وهذا مثال آخر على استخدام قانون حرية المعلومات في البحث، فمن الوسائل التي ساعدتنا بصورة كبيرة أننا أخبرنا السلطات الهولندية أن البيانات متاحة في الدنمارك.

ولكن الخبر كان صحيحًا: ففي أوروبا كان هناك شباب تملكتهم النزعة الانتحارية وللأسف كانت هناك حالات انتحار أيضًا في عدة بلدان نتيجة لتناول العقار. قام الصحافيون والباحثون وعائلة أحد الضحايا بالضغط للحصول على تلك المعلومات. كما ساعد أمين المظالم الأوروبي في الضغط من أجل الشفافية في وكالة الأدوية الأوروبية، ويبدو أنه نجح في ذلك (http: / /bit.ly /eu-ombuds-man). إذن فالآن يقع على عاتق الصحافيين مهمة استخراج البيانات وتحليلها بدقة. هل جميعنا فئران تجارب كما قال أحد الباحثين؟ أم أن آليات المراقبة سليمة؟

الدرس المستفاد: لا تقبل الرفض عندما يتعلق الأمر بالشفافية. بل كُن مثابرًا وتتبع الخبر مع مرور الوقت. فقد تتغير الأمور وتسمح بإعداد تقارير أفضل بناء على فرص أفضل تتيح الحصول على البيانات في مرحلة لاحقة.

دراسة حالة رقم (3): تهريب الموت

قد يكون التاريخ القريب مؤلمًا للغاية بالنسبة لشعوب بأكملها، خاصة بعد الحروب وأثناء الفترات الانتقالية. كيف إذن يمكن للصحافيين أن يحصلوا على بيانات صعبة للتحقيق عندما يكون المستفيدون من حرب العقد الأخير– مثلاً – على رأس السلطة في الوقت الحاضر؟ كانت تلك هي المهمة التي شرع في تحقيقها

فريق من الصحافيين السلوفانيين والكروات والبوسنيين.

شرع الفريق في التحقيق في تجارة الأسلحة في يوغوسلافيا السابقة أثناء فترة الحظر التي فرضتها الأمم المتحدة في بداية التسعينيات. كان العمل مبنيًا على وثائق من التحقيقات البرلمانية في هذا الصدد. ومن أجل توثيق مسارات الشحن وفهم هيكل التجارة، كان يجب تتبع عمليات النقل من خلال أرقام الحاويات في الموانئ ولوحات تسجيل الشاحنات.

أجرت لجان البرلمان السلوفينية تحقيقات في قضية التربّح من حروب البلقان، ولكنها لم تتوصل إلى أي نتيجة. ومع ذلك، كان هناك أثر في غاية الأهمية من الوثائق والبيانات التي رُفعت عنها السرية، بما في ذلك 6000 صفحة حصل عليها الفريق السلوفيني من خلال تقديم طلب حرية المعلومات.

في تلك الحالة، كان يجب استخراج البيانات من الوثائق وتخزينها في قواعد بيانات. ومع إضافة المزيد من البيانات وإجراء التحليل والبحث، تمكّن الفريق من رسم خريطة للمسارات العديدة لتجارة الأسلحة غير القانونية (http://bit.ly/kaasogmulvad-smuggling).

نجح الفريق وكانت النتائج فريدة (http://bit.ly/journalismfund-smuggling1) وقد حاز الفريق على جائزته الأولى. والأهم من ذلك أن الخبر يهم المنطقة بأكملها وقد يتناوله صحافيون في البلدان الأخرى التي مرت من خلالها الشحنات القاتلة.

الدرس المستفاد: استخرج مواد خام أولية جيدة حتى وإن وجدتها في أماكن غير متوقعة ثم اجمعها مع البيانات المتاحة للعامة بالفعل.

--- بريجيت ألفتر، موقع Journalismfund.eu

استخدم قانون حرية المعلومات مع الأصدقاء

لدى العديد من دول البلقان مشاكل مع فساد الحكومات. فالفساد عادة ما يكون أكبر فيما يتعلق بمُساءلة الحكومات المحلية في تلك الدول. على مدى عدة أشهر، قامت مجموعة من الصحافيين الصرب في مركز الصحافة الاستقصائية في بلغراد (http://www.cins.org.rs/) بالتحقيق في أنواع مختلفة من وثائق

قانون حرية المعلومات من أكثر من 30 بلدية محلية عام 2009. وقبل ذلك، لم يكن هناك أي شيء تقريبًا متاحاً للجمهور. كانت الفكرة تكمن في الحصول على السجلات الحكومية الأصلية ووضع البيانات في جداول بيانية وإجراء فحوص ومقارنات أساسية بين البلديات والحصول على أعلى وأدنى الأرقام. كانت المؤشرات الأساسية هي أرقام الموازنة والنفقات العادية والخاصة ومرتبات المسؤولين ونفقات السفر وأعداد الموظفين ونفقات الهواتف المحمولة والبدلات اليومية وأرقام المشتريات العامة وما إلى ذلك. وقد كانت تلك هي المرة الأولى التي يطلب فيها الصحافيون مثل تلك المعلومات.

كانت النتيجة قاعدة بيانات شاملة تكشف عن عدة نتائج زائفة وتجاوزات وحالات فساد. وقد أشارت لائحة رؤساء البلديات الأعلى أجرًا إلى أن بعضهم يتلقى أموالاً أكثر من الرئيس الصربي. كان هناك العديد من المسؤولين الآخرين الذين يتقاضون رواتب زائدة، وهناك أيضًا العديد ممن يتقاضون بدلات سفر وبدلات يومية هائلة. وقد ساعدتنا بيانات المشتريات العامة التي بذلنا مجهودًا للحصول عليها في تسليط الضوء على فوضى رسمية. فقد صدر عن قاعدة البيانات تلك أكثر من 150 خبرا وقامت وسائل الإعلام المحلية والوطنية في صربيا بنشر العديد من تلك الأخبار.

تعلمنا أن مقارنة السجلات مع البيانات المماثلة من جهات حكومية مشابهة يمكنها أن تظهر الانحرافات وأن تسلط الضوء على الفساد المحتمل. يمكن الكشف عن النفقات المبالغ فيها والمصروفات غير العادية فقط عن طريق المقارنة.

— جورجي باديسكي، عضو زمالة Knight Journalism في جامعة ستانفورد

الحصول على البيانات من الويب

هل جربتَ كل شيء آخر ولم تتمكن من الحصول على البيانات التي تريدها؟ هل وجدتَ البيانات على الإنترنت، ولكن لا توجد خيارات للتحميل وقد خذلك خيار النسخ؟ لا تقلق، لا تزال هناك وسيلة للحصول على البيانات، فيمكنك على سبيل المثال:

- الحصول على البيانات من وصلات تداخل لبرمجة التطبيقات API على الإنترنت، فهذه الوصلات توفرها العديد من قواعد البيانات وتطبيقات الإنترنت الحديثة على الشبكة (بما فيها تويتر وفيسبوك والعديد من التطبيقات الأخرى). هذه طريقة رائعة للحصول على البيانات الحكومية أو التجارية، بالإضافة إلى البيانات من مواقع الإعلام الاجتماعي.
- استخراج البيانات من ملفات PDF مسألة صعبة للغاية، فلغة PDF خاصة بالطابعات ولا تحفظ الكثير من المعلومات بنفس صورة هيكل البيانات الموجودة داخل الوثيقة. واستخراج البيانات من ملفات PDF ليس من اختصاصات هذا الكتاب، ولكن هناك بعض الأدوات والدروس التوضيحية التي قد تساعدك في ذلك.
- استخراج البيانات من خلال التفقد الدقيق للمواقع الإلكترونية. أثناء قيامك بذلك، أنت تستخرج محتوى منظم من صفحة إلكترونية عادية بمساعدة إحدى أدوات البحث واستخراج البيانات أو بكتابة بعض الرموز. وعلى الرغم من أن هذه الطريقة قوية جدًا ويمكن استخدامها في عدة أماكن، إلا أنها تتطلب بعض الفهم لكيفية عمل الويب.

مع كل تلك الخيارات التقنية الرائعة، لا تنس الخيارات البسيطة: ففي الكثير من الأحيان من الجدير أن تقضي بعض الوقت في البحث عن ملف يحتوي على بيانات مقروءة آليًا أو أن تتصل بالمؤسسة التي تحتفظ بالبيانات التي تريدها.

في هذا الفصل سنأخذ جولة لاستعراض مثال بسيط جدًا لاستخراج البيانات من صفحة ويب HTML.

ما هي البيانات المقروءة آليًا؟

معظم هذه الأساليب تهدف إلى الحصول على بيانات مقروءة آليًا، وهي التي يتم إنشاؤها ليعالجها الكمبيوتر بدلاً من المعالجة البشرية. ويعتمد هيكل بناء البيانات على ما تحتويه من معلومات، وليس على طريقة عرضها النهائية. هناك أمثلة للصيغ التي تسهل قراءتها آليًا ومنها ملفات CSV وXML وJSON وExcel، بينما هناك صيغ مثل وثائق Word وصفحات HTML وملفات PDF تهتم أكثر

بالعرض المرئي للمعلومات. فمثلاً، PDF هي لغة تخاطب طابعتك مباشرة، وتهتم بموقع السطور والنقاط على الصفحة.

ما الغرض من استخراج البيانات من المواقع الإلكترونية؟

لقد تعرض الجميع إلى: الذهاب إلى موقع إلكتروني، وعند رؤية جدول مثير للاهتمام نحاول نسخه إلى إكسل حتى نتمكن من جمع بعض الأرقام أو تخزينها لوقت لاحق. ولكن هذا الأمر لا يفلح في الغالب، أو أن المعلومات التي نريدها متفرقة بين عدد كبير من المواقع الإلكترونية. وقد يتحول النسخ اليدوي سريعًا إلى أمر ممل للغاية، إذن فمن المنطقي أن نستخدم القليل من الرموز لتقوم بذلك بالنيابة عنا.

تعد ميزة استخراج البيانات من المواقع الإلكترونية في إمكانية القيام بذلك بأي موقع إلكتروني تقريبًا، بدءًا من النشرات الجوية وصولاً إلى الإنفاق الحكومي، حتى وإن لم يكن لدى هذا الموقع وصلة تداخل لبرمجة التطبيقات API للحصول على البيانات الخام.

ما يمكن وما لا يمكن استخراجه من البيانات

عملية استخراج البيانات لها حدود. ومن ضمن العوامل التي تزيد من صعوبة استخراج البيانات من المواقع الإلكترونية:

- الصياغة السيئة لرمز HTML مع وجود القليل من المعلومات المنظمة – أو انعدامها – مثل المواقع الحكومية القديمة.
- أنظمة التوثيق التي من المفترض أن تمنع الدخول التلقائي، مثل رمز -CAPT CHA أو paywalls الذي يمنع الدخول إلى المواقع قبل تسديد الرسوم.
- الأنظمة القائمة على الجلسات، والتي تستخدم ملفات cookies لتتبع ما يفعله المستخدم.
- نقص القوائم الكاملة واحتمالات البحث بحروف بديلة Wildcard.
- منع الدخول الجماعي للموقع من قبل مسؤولي مزود خدمات الويب.

هناك أيضًا مجموعة أخرى من العوائق وهي الحواجز القانونية: تعترف بعض البلدان بحقوق التأليف والنشر لقواعد البيانات، مما قد يحدّ من حقك في إعادة

استخدام المعلومات التي سبق نشرها على الإنترنت. وأحيانًا يمكنك تجاهل الحصول على إذن والقيام بالنشر على أية حال، وهذا يعتمد على صلاحياتك، فقد تكون لديك حقوق خاصة بصفتك صحافيًا بحيث تمر عملية استخراج البيانات الحكومية المتاحة بدون مشاكل، ولكن يُفضل أن تتأكد قبل النشر. أما المنظمات التجارية وبعض المنظمات غير الحكومية المعينة، فقد تتعامل بقدر قليل من التسامح وقد تحاول الادعاء بأنك 'تخرب' أنظمتها. وقد يتسبب استخراج البيانات من معلومات أخرى في انتهاك خصوصية الأفراد وبالتالي تخرق قوانين خصوصية البيانات أو أخلاقيات المهنة.

ترقيع البيانات واستخراجها وتجميعها وتنقيتها

لا يكمن تحدي التعامل مع كميات هائلة من بيانات المملكة المتحدة في نشرها، بل في وضعها في صيغة قابلة للاستخدام. فالعديد من البيانات حول الضيافة، والمصالح الخارجية للنواب، وجماعات الضغط يتم نشرها بانتظام ولكن بطرق يصعب تحليلها.

بالنسبة لبعض المعلومات، لا يوجد إلا الطريق الصعب: ترقيع العشرات من ملفات إكسل التي يحتوي كل منها على بضع عشرات فقط من السجلات، وكانت تلك هي الطريقة الوحيدة لوضع قوائم شاملة للاجتماعات الوزارية. أما بالنسبة لبعض المعلومات الأخرى، فقد ثبت أن استخراج البيانات من الويب أمر مفيد للغاية.

وقد أنجزنا تقريبًا نصف عملنا من خلال استخدام خدمة مثل ScraperWiki لنطلب من المبرمجين صنع برنامج لاستخراج المعلومات مثل سجل مصالح النواب، فأصبحت لدينا كل معلومات النواب في صفحة واحدة وجاهزة لمهمة التحليل والتنقية الطويلة.

إن مثل تلك الخدمات (أو الأدوات مثل Outwit Hub) تمثل مساعدة هائلة للصحافيين الذين يحاولون جمع البيانات الفوضوية والذين لا يستطيعون كتابة الرمز بأنفسهم.

— جيمس بال، صحيفة الغارديان

الأدوات المساعدة على استخراج البيانات

هناك العديد من البرامج التي يمكن استخدامها لاستخراج كتل المعلومات من المواقع الإلكترونية، منها ملحقات المتصفح وبعض خدمات الويب. حسب المتصفح الذي تستخدمه، فإن الأدوات مثل Readability (http://www.readability.com/) (التي تساعد في استخراج النص من الصفحة) أو Down-ThemAll (http://www.downthemall.net/) (التي تتيح لك تحميل عدة ملفات في نفس الوقت) ستساعدك في جعل بعض المهام المملة آلية، ولقد تم تطوير ملحق Chrome Scraper (http://bit.ly/chrome-scraper) خصيصا لاستخراج الجداول من المواقع الإلكترونية. وتسمح لك ملحقات المطور مثل FireBug (http://getfirebug.com/) (وهذا لمتصفح Firefox، فالأمر ذاته موجود في Chrome وSafari وIE) بتتبع هيكلة الموقع الإلكتروني وأي الاتصالات تحدث بين متصفحك وبين خادم الويب.

يسمح لك موقع ScraperWiki (https://scraperwiki.com) بكتابة الرمز لأدوات استخراج البيانات بعدد من مختلف لغات البرمجة، بما فيها بايثون Python وروبيRuby وPHP. إذا أردت أن تبدأ في استخراج البيانات دون معاناة تحضير بيئة برمجية على الكمبيوتر الخاص بك، فهذه هي الطريقة. هناك خدمات ويب أخرى مثل جداول جوجل البيانية Google Spreadsheets وYahoo! Pipes التي تتيح لك استخراج بعض البيانات من مواقع أخرى كذلك.

كيفية عمل أداة استخراج البيانات من الويب

عادة ما تكون أدوات استخراج البيانات من الويب عبارة عن أجزاء صغيرة من الرمز مكتوبة بلغة برمجة مثل بايثون أو روبي أو PHP. واختيار اللغة المناسبة يعتمد في الغالب على المجتمع الذي تتواصل معه: فإذا كان لديك أحد في غرفة الأخبار أو في مدينتك يعمل بالفعل باستخدام إحدى تلك اللغات، فمن المنطقي أن تستخدم نفس اللغة.

بعض أدوات استخراج البيانات 'بالنقر والتوجيه' التي سبق ذكرها قد تفيد كبداية، ولكن يكمن التعقيد الحقيقي المتعلق باستخراج البيانات من الويب

في التعامل مع الصفحات والعناصر المناسبة داخل تلك الصفحات لاستخراج المعلومات المنشودة. ولا تتعلق هذه المهام بالبرمجة، بل بفهم هيكل الموقع الإلكتروني وقاعدة البيانات.

عند عرض الموقع الإلكتروني، سيحاول المتصفح دائمًا أن يستفيد تقريبًا من تقنيتين: HTTP وهي طريقة للتواصل مع خادم الويب "سيرفر"، ولطلب مورد محدد، مثل الوثائق أو الصور أو مقاطع الفيديو، وHTML وهي اللغة التي تتكون منها المواقع الإلكترونية.

تشريح صفحة الويب

أي صفحة HTML تتألف من تسلسل هرمي مكون من مربعات (تحددها "علاماتHTML "tags). يحتوي كل مربع كبير على عدة مربعات أصغر حجمًا – مثلًا جدول به العديد من الأقسام الأصغر حجما: صفوف وخانات. وتوجد عدة أنواع من العلامات التي تؤدي وظائف مختلفة – فبعضها ينشئ مربعات، والبعض الآخر ينشئ جداول أو صورًا أو روابط. يمكن للعلامات أيضًا أن يكون لديها خصائص إضافية (يمكنها مثلاً أن تكون وسائل فريدة لتحديد الهوية) ويمكن أن تنتمي إلى مجموعات تُسمى 'تصنيفات'، مما يجعل من الممكن استهداف عناصر فردية وعرضها داخل وثيقة. يذكر أن اختيار العنصر الملائم بهذه الطريقة واستخراج محتواه يعتبر مفتاحًا لاستخراج البيانات.

بالنظر إلى عناصر صفحة الويب، يمكن تقسيم كل شيء إلى مربعات داخل مربعات.

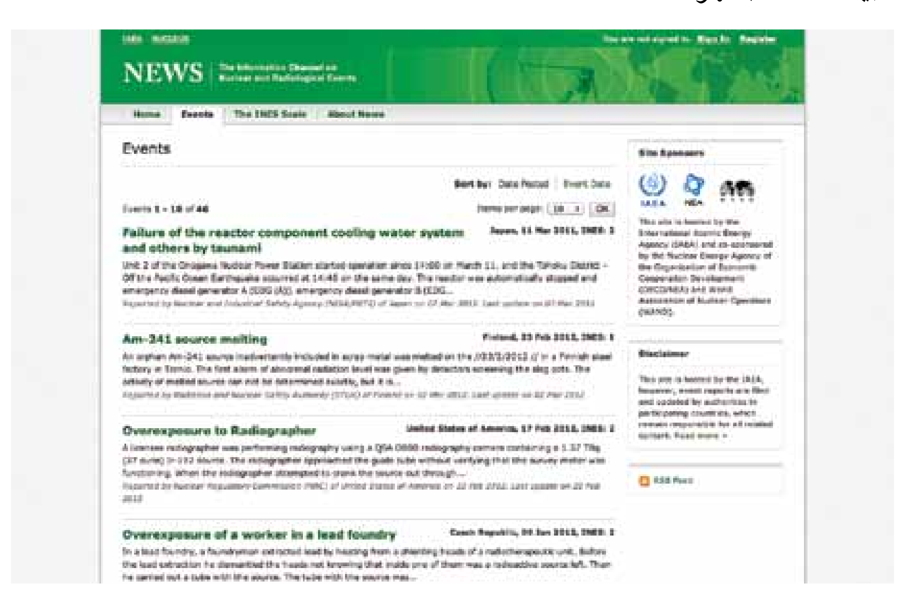
حتى تستخرج البيانات من صفحات الويب، ستحتاج إلى التعرف قليلاً على الأنواع المختلفة للعناصر التي يمكن أن تكون في وثيقة HTML. فعلى سبيل المثال، يشمل عنصر <table> الجدول بالكامل، والذي يحتوي على عناصر <tr> (صف الجدول table row) للصفوف، والتي تحتوي بدورها على عنصر <td> (بيانات الجدول table data) لكل خانة. إن أكثر أنواع العناصر رواجاً هو <div> والذي يعني أساسًا أي كتلة من المحتوى. وأسهل طريقة لفهم تلك العناصر هي استخدام شريط أدوات المطوّر (http://bit.ly/developer-toolbar) في

متصفحك: فسيسمح لك بالوقوف بالمؤشر فوق أي جزء من صفحة الويب ورؤية الكود الذي تتكون منه.

تعمل العلامات مثل مساند الكتب، فهي تحدد بداية ونهاية أي وحدة. فمثلاً، تشير علامة إلى بداية نص ذي حروف مائلة أو محددة وتشير إلى نهاية هذا الجزء. إذن فالأمر سهل.

مثال: استخراج بيانات الحوادث النووية باستخدام لغة بايثون

يشتمل موقع NEWS (http://www-news.iaea.org/EventList.aspx) وهو بوابة 'الوكالة الدولية للطاقة الذرية' (IAEA) على حوادث الإشعاع حول العالم (كما أنه ينافس بقوة على عضوية 'نادي العناوين الغريبة'!). تنشر الصفحة قوائم بالحوادث في موقع بسيط يشبه المدونات، ويمكن استخراج البيانات منه بسهولة.



نموذج 4–4. بوابة 'الوكالة الدولية للطاقة الذرية' (news.iaea.org) (IAEA)

للبدء قم بإنشاء أداة بايثون جديدة لاستخراج البيانات على موقع Scraper-Wiki (https://scraperwiki.com/) وستحصل على مساحة نصية أغلبها خال، فيما عدا بعض أنواع الرمز الداعم. في نافذة متصفح أخرى، افتح موقع IAEA

(http://www-news.iaea.org/EventList.aspx) وافتح شريط أدوات المطوّر في متصفحك. في خانة 'العناصر'، حاول أن تجد عنصر HTML لأحد عناوين الخبر. يساعدك شريط أدوات المطوّر في متصفحك على توصيل العناصر في صفحة الويب بكود HTML الذي تتكون منه.

سيكشف لك التحقق من تلك الصفحة أن العناوين في صيغة عناصر <h4> داخل الجدول <table>. كل حدث يتكون من صف <tr>، والذي يحتوي أيضًا على وصف وتاريخ. وإذا أردنا استخراج عناوين كل الأحداث، علينا أن نجد طريقة لاختيار كل صف في الجدول بالتعاقب، بينما نجلب كل النص داخل عناصر العنوان.

لتحويل تلك العملية إلى الرمز، علينا أن ندرك كل الخطوات المعنية. ولأخذ فكرة عن نوع الخطوات المطلوبة، هيا بنا نلعب لعبة بسيطة: في نافذة ScraperWiki لديك، حاول أن تكتب إرشادات أو تعليمات فردية لنفسك تتضمن كل شيء ستفعله أثناء كتابة أداة استخراج تلك البيانات، مثل خطوات الطبخ في وصفة الطعام (ابدأ كل سطر بعلامة تجزئة # لتخبر بايثون بأن هذا ليس رمز كمبيوتر حقيقيًّا). على سبيل المثال:

```
# Look for all rows in the table
# Unicorn must not overflow on left side.
```

حاول أن تكون محددًا بقدر الإمكان ولا تفترض أن البرنامج يعلم أي شيء عن الصفحة التي تحاول استخراج البيانات منها.

بمجرد أن تكتب بعض الرمز الافتراضي، دعنا نقارنه بالرمز الأساسي لأول أداة استخراج بيانات:

```
import scraperwiki
from lxml import html
```

في هذا القسم الأول، نقوم باستيراد وظائف تشغيل موجودة بالفعل من المكتبات – مقتطفات من الرمز المكتوب مسبقًا. سيعطينا scraperwiki القدرة على تحميل مواقع الويب، بينما يعتبر lxml أداة للتحليل المنظم لوثائق HTML. هناك خبر جيد، وهو أنك إن كنت تكتب أداة استخراج بيانات بلغة بايثون مستخدما ScraperWiki، سيظل هذان السطران على نفس الشكل.

```
"url = "http://www–news.iaea.org/EventList.aspx
(doc_text = scraperwiki.scrape(url
(doc = html.fromstring(doc_text
```

بعد ذلك يكوّن الرمز اسمًا (متغير): عنوان url، ويحدد عنوان URL الخاص بـ IAEA كقيمة خاصة به. وهذا يخبر أداة استخراج البيانات أن هذا الشيء موجود ونود أن نعيره اهتمامنا. ولاحظ أن عنوان URL نفسه موجود بين علامتي اقتباس لأنه ليس جزءًا من رمز البرنامج ولكنه سلسلة (String)، أي سلسلة محارف (حروف وأشكال).

نقوم بعد ذلك باستخدام متغير عنوان url كمدخل بالدالة، scraperwiki. scrape. وعليه ستقدم الدالة دورًا محددًا، وفي تلك الحالة ستقوم بتحميل صفحة الويب. وعند الانتهاء من ذلك، ستضع المُخرَج في متغير آخر وهو doc_text الذي سيحمّل النص الفعلي للموقع – ليس بالشكل المرئي الذي تراه في متصفحك، ولكن بالرمز المصدري وكل ما فيه من علامات. وبما أن تحليل هذا الشكل ليس سهلاً، فسنستخدم دالة أخرى وهي html.fromstring لتوليد تمثيل خاص حيث يمكننا بسهولة معالجة العناصر وما يُسمى بـ 'نموذج كائن المستند ' (DOM).

```
:("for row in doc.cssselect("#tblEvents tr
()link_in_header = row.cssselect("h4 a").pop
event_title = link_in_header.text
print event_title
```

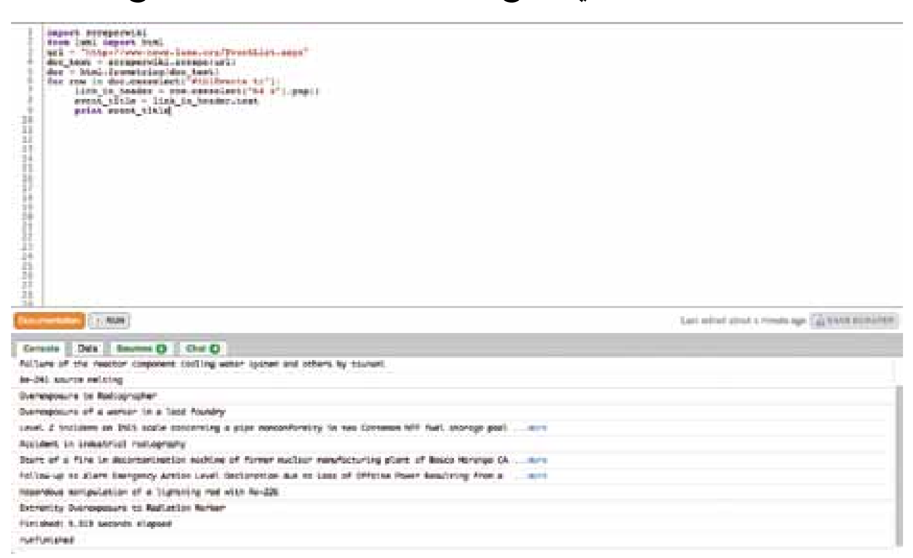
في تلك الخطوة الأخيرة، نستخدم DOM لنجد كل صف في جدولنا ونستخرج عنوان الحدث من رأس الصفحة. ويتم استخدام مفهومين جديدين: (for loop) وعنصر الاختيار (.cssselect). تقوم for loop أساسًا بما يوحي به اسمها، أي أنها تجتاز قائمة من العناصر وتعطي كلا منها رمزًا مكافئًا مؤقتًا (في تلك الحالة: الصف) ثم تقوم بتنفيذ أي تعليمات خاصة بكل عنصر.

أما المفهوم الجديد الآخر، اختيار العنصر، فيستخدم لغة خاصة ليجد العناصر في الوثيقة. عادة ما تُستخدم أدوات اختيار CSS لإضافة معلومات التصميم إلى عناصر HTML ويمكن استخدامها لانتقاء عنصر محدد من الصفحة. في تلك

الحالة (السطر السادس)، نقوم باختيار tblEvents tr# مما يتطابق مع كل عنصر <tr> داخل عناصر الجدول بهوية tblEvents (فعلامة التجزئة # تشير ببساطة إلى الهوية). لاحظ أن هذا سينتج عنه قائمة بعناصر <tr>.

وكما ترى في السطر التالي (السطر السابع)، حيث نضع أداة اختيار أخرى لنجد أي عنصر <a> (الذي يعني رابط) داخل عنصر <h4> (عنوان). هنا نريد فقط أن ننظر في عنصر واحد (فهناك عنوان واحد فقط في كل صف)، لذا علينا أن نحركه من على أعلى القائمة التي أعادتها إلينا أداة الاختيار باستخدام دالة ()pop.

لاحظ أن بعض العناصر في DOM تحتوي نصًا فعليًا، (أي نصًا ليس جزءًا من أي لغة ترميز)، ويمكننا الدخول إليه باستخدام تركيب element].text] الذي نراه في السطر الثامن. وأخيرًا، في السطر التاسع، نطبع هذا النص في لوحة تحكم ScraperWiki. وإذا ضغطت على زر التشغيل في أداة استخراج البيانات لديك، ستبدأ النافذة الأصغر الآن في وضع قائمة بأسماء الأحداث من موقع IAEA.



نموذج 4–5. أداة استخراج البيانات أثناء عملها (ScraperWiki)

يمكنك الآن رؤية أداة استخراج بيانات أساسية أثناء عملها: فهي تقوم بتحميل صفحة الويب وتحولها إلى صيغة DOM ثم تسمح لك بانتقاء واستخراج محتوى معين. وبهذا الهيكل، يمكنك أن تحاول حل بعض المشاكل المتبقية باستخدام

توثيق ScraperWiki وبايثون:

- هل تستطيع العثور على عنوان الرابط في العنوان الرئيسي كل حدث؟
- هل تستطيع اختيار المربع الصغير الذي يحتوي على التاريخ والمكان باستخدام تصنيف CSS الخاص به واستخراج نص العنصر؟
- يوفر ScraperWiki قاعدة بيانات صغيرة لكل أداة استخراج بيانات لتتمكن من تخزين النتائج ونسخ المثال ذي الصلة من الوثائق وتهيئته حتى يحفظ عناوين الأحداث والروابط والتواريخ.
- تتكون قائمة الأحداث من عدة صفحات، هل تستطيع استخراج البيانات من عدة صفحات للحصول على الأحداث التاريخية أيضًا؟

وبينما تحاول حل هذه التحديات، الق نظرة على ScraperWiki: فهناك العديد من الأمثلة المفيدة في أدوات استخراج البيانات الموجودة بالفعل، وفي كثير من الأحيان تكون البيانات شائقة جدًا أيضًا. هكذا لا تحتاج إلى بدء أداتك من الصفر: فقط اختر أداة مشابهة وعدلها وهيئها لتناسب بحثك.

— فريدريك ليندنبرغ، مؤسسة المعرفة المفتوحة

استخراج البيانات من قاعدة بيانات عامة

هناك بعض الأطباء الفرنسيين الذين لديهم حرية تحديد رسوم الكشف الخاصة بهم، فيمكن للمرء مثلاً أن يدفع ما بين 70 إلى 500 يورو لقاء زيارة مدتها 30 دقيقة لطبيب الأورام. هذه المعلومات الخاصة بالرسوم متاحة للجمهور بموجب القانون، ولكن الإدارة توفر فقط قواعد البيانات الإلكترونية التي يصعب تصفحها. ولأتمكن من إلقاء نظرة متفحصة على رسوم كشف الأطباء لصالح صحيفة لوموند، قررت أن أستخرج البيانات من قاعدة البيانات بأكملها.

وهنا بدأت الإثارة. كان نموذج البحث المبدئي عبارة عن تطبيق فلاش يقوم بإعادة التوجيه إلى صفحة HTML للنتائج عن طريق طلب POST. وبمساعدة نيكولاس كايزر–بريل، استغرقنا بعض الوقت لندرك أن التطبيق استخدم صفحة اضافية ثالثة كخطوة 'خفية' بين نموذج البحث وصفحة النتائج. استخدمت هذه

الصفحة لتخزين ملف تعريف ارتباط (cookie) به قيم من نموذج البحث الذي تمكنا بعد ذلك من الوصول إليه من خلال صفحة النتائج. كان من الصعب التفكير في عملية أكثر تعقيداً، ولكن خيارات مكتبة cURL بلغة PHP تسهل تجاوز العقبات بمجرد أن تعرف مكانها! في النهاية، استغرقنا 10 ساعات للحصول على قاعدة البيانات، ولكن المهمة كانت تستحق ذلك.

—— أليكساندر ليشنيت، صحيفة لوموند

الويب كمصدر للبيانات

كيف يمكنك معرفة المزيد عن شيء لا يوجد إلا على الإنترنت؟ سواء أكنت تبحث عن عنوان بريد إلكتروني أو موقع أو صورة أو مقالة في ويكيبيديا، سأطلعك في هذا الفصل على الأدوات التي ستخبرك بالمزيد عن خلفياتها.

أدوات الويب

أولاً، هناك بعض الخدمات المختلفة التي يمكنك الاستفادة منها لاكتشاف المزيد عن موقع إلكتروني بأكمله، بدلًا من صفحة واحدة معينة.

موقع Whois

إذا فتحت صفحة http://whois.domaintools.com/ (أو فقط اكتب whois www.example.com في تطبيق Terminal.app على جهاز ماكنتوش Mac،) يمكنك الحصول على معلومات التسجيل الأساسية لأي موقع. في السنوات الأخيرة، قام بعض المُلاك باختيار التسجيل 'الخاص' الذي يخفي تفاصيلهم، ولكن في الكثير من الحالات ستجد الاسم والعنوان والبريد الإلكتروني ورقم الهاتف للشخص الذي قام بتسجيل الموقع. كما يمكنك كذلك إدخال عناوين IP رقمية وستحصل على بيانات المنظمة أو الشخص الذي يملك خادم الويب (السيرفر). ويكون ذلك مفيداً على وجه الخصوص حين تحاول تتبع المزيد من المعلومات عن مستخدم مؤذ أو مهاجم لخدمة معينة، حيث إن معظم المواقع تسجل عنوان IP لجميع زوارها.

محرك البحث Blekko

يقدم محرك بحث Blekko (/http://blekko.com) كماً استثنائيا من الأفكار حول الإحصاءات الداخلية التي يجمعها عن المواقع أثناء زحفه عبر الشبكة. إذا كتبت اسم مجال domain وأتبعته بـ '/seo' ستتلقى صفحة معلومات حول هذا العنوان توضح لك النافذة الأولى في نموذج 4–7 أي مواقع أخرى ترتبط بالمجال مُرتَّبة حسب رواجها. وقد يكون ذلك مفيدًا للغاية حين تحاول فهم التغطية التي يتلقاها موقع ما، وإذا أردت أن تفهم لماذا يظهر في ترتيب عالٍ في نتائج بحث جوجل، بما أنها مبنية على تلك الروابط الواردة. كما يخبرك نموذج 4–8 أي المواقع الأخرى التي تشغلها نفس الآلة. فمن الشائع بين المحتالين والمتطفلين أن يستخدموا الدعاية ليحققوا شرعية مصطنعة عن طريق بناء مواقع عديدة متصلة ببعضها بعضًا. ولكن تبدو وكأن مجالاتها مستقلة، وقد تكون لديها تفاصيل تسجيل مختلفة، ولكنها عادة ما تكون على نفس خادم الويب، وبذلك تقل التكلفة كثيرا. تلك الإحصاءات تعطيك فكرة عن الهيكل التجاري الخفي للموقع الذي تبحث عنه.

blekko beta
petewarden.typepad.com /seo search
examples: cure for headaches | global warming /liberal

نموذج 4–6. محرك البحث Blekko (Blekko.com)

Inbound links: 6,050 from 302 domains:

#	from host	host rank	links	last	actions
1	twitter.com	12,366.4	1		
2	www.guardian.co.uk	6,481.2	1		
3	www.forbes.com	3,699.8	1	41d ago	
4	www.newscientist.com	3,678.4	2		
5	code.google.com	3,451.1	1		
6	www.huffingtonpost.com	3,238.2	1		
7	news.cnet.com	3,185.8	2		
8	gizmodo.com	2,119.3	6	39d ago	

نموذج 4–7. فهم الرواج على الويب: من يرتبط بمن؟ النافذة المفيدة الأخرى هي 'إحصائيات الزحف

‘Crawl stat وخاصة الجزء الخاص بـ ’المضيف المشارك’ (Blekko.com ’Cohosted with)

Cohosted With:

host	whois	view
thelongtail.com	whois	
codinghorror.com	whois	
longtail.com	whois	
cityofsound.com	whois	
hypebot.com	whois	
therestisnoise.com	whois	
stevenberlinjohnson.com	whois	
planetout.com	whois	
riehlworldview.com	whois	

نموذج 4–8. مراقبة المحتالين والمتطفلين (Blekko.com)

موقع Compete.com

من خلال دراسة شريحة من المستهلكين الأمريكيين، يراكم موقع Com- pete.com (/http: / /www.compete.com) إحصاءات مفصلة عن استخدام معظم المواقع، ويتيح بعض التفاصيل الأساسية بالمجان. اختر ‘لمحة عن الموقع’ وأدخل اسم المجال (نموذج 4–9) وسيظهر لك رسمًا بيانيًا لزيارات الموقع خلال العام الماضي، بالإضافة إلى عدد الزائرين والزيارات (نموذج 4–10). وبما أن الأرقام مبنية على الدراسات، فهي تقريبية فقط، ولكنني وجدتها دقيقة عندما قارنتها مع التحليلات الداخلية، وخاصة أنها تشكل مصدرًا جيدًا عند مقارنة موقعين لأنه مع غياب الأرقام المطلقة لكلا الموقعين، تظل تلك الأرقام تقدم تمثيلا جيدًا للفارق النسبي في رواجهما. ولكن هذا الموقع يجري دراسات عن المستهلكين الأمريكيين فقط، لذا فالبيانات المتوفرة عن المواقع ذات الزيارات الدولية الكثيرة سوف تكون ضعيفة.



نموذج 4–9.خدمة ’لمحة عن الموقع‘ من موقع Compete.com (Compete.com)



نموذج 4–10. ما هو الرائج؟ ما هو المرغوب؟: مواقع الويب الرائجة (Compete.com)

خدمة جوجل للبحث عن المواقع

هناك خاصية مفيدة للغاية حين تحاول استكشاف المحتوى الكامل لمجال معين وهي الكلمة البحثية 'موقع site:'. فإذا أضفت 'site:example.com' إلى جملتك البحثية، سيعرض لك جوجل النتائج من الموقع الذي حددته فقط. يمكنك حتى أن تجعل البحث أكثر تحديداً من خلال إضافة بادئات الصفحات التي تهمك، مثلاً اكتب 'site:example.com /pages/ ' وستحصل فقط على النتائج المطابقة لهذا النمط. قد يفيدك ذلك للغاية حين تحاول البحث عن المعلومات التي أتاحها مُلاك المجال للجمهور ولكنهم لا يريدون الإعلان عنها، لذا فانتقاء الكلمات البحثية المناسبة يمكن أن يفشي مواد كاشفة للغاية.

صفحات الويب والصور ومقاطع الفيديو

أحيانًا تكون مهتمًا بالنشاط الذي يحيط بخبر معين وليس بالموقع بأكمله، الأدوات الآتية ستطلعك على زوايا مختلفة حول كيفية قراءة الناس لمحتوى الويب واستجابتهم ونسخهم وتعميمهم لهذا المحتوى.

موقع Bit.ly

حين أرغب في معرفة كيف يعمم الناس رابطًا معينًا بين بعضهم البعض، فدائمًا ما ألجأ إلى موقع bit.ly (http://bitly.com/). ولاستخدام الموقع، أدخل عنوان URL الذي يهمك، ثم اختر رابط 'InfoPage+'. سيأخذك هذا إلى صفحة

الإحصاءات الكاملة (ولكنك قد تحتاج إلى اختيار 'aggregate bit.ly link' أولاً إذا كنت مشتركًا في الخدمة). وسيعطيك ذلك فكرة عن مدى رواج الصفحة، بما في ذلك نشاطها على موقعي فيسبوك وتويتر، وفي الأسفل سترى محادثات عامة حول الرابط يوفرها موقع backtype.com. أجد هذا الجمع بين بيانات الزيارات والمحادثات مفيدًا للغاية حين أحاول فهم السبب وراء رواج موقع أو صفحة ما، ومن هو الجمهور الذي يزوره؟ فمثلاً، زودني هذا بدليل قاطع على خطأ الرواية السائدة بشأن المشاركة الشعبية مع "سارة بالين".

موقع تويتر

مع اتساع رقعة خدمة التدوين المصغر، أصبحت الخدمة مفيدة كمعيار لمشاركة الناس وتحدثهم عن محتوى معين. ويمكنك استكشاف المحادثات العامة بشأن رابط ما باستخدام خدعة بسيطة. فقط انسخ عنوان URL الذي يهمك في خانة البحث ثم اختر 'المزيد من التغريدات' لترى النتائج الكاملة.

ذاكرة التخزين المؤقت Cache من جوجل

عندما تصبح صفحة ما مثيرة للجدل، قد يزيلها الناشر أو يعدلها دون أي إشعار. وإذا كنت تشك في أنك تواجه تلك المشكلة، فإن أول ما تلجأ إليه هو ذاكرة التخزين المؤقت من جوجل للصفحة كما كانت عندما تم الزحف عليها في المرة الأولى. ومن الجدير بالذكر أن عدد مرات الزحف تزداد باستمرار، لذا سيحالفك الحظ كثيرًا إذا حاولت القيام بذلك في الساعات الأولى من التغييرات التي تشك في حدوثها. أدخل عنوان URL المنشود في خانة جوجل للبحث، ثم اختر السهم الثلاثي على يمين نتيجة تلك الصفحة. سيظهر لك عرض رسومي مسبقًا، وإذا كنت محظوظًا، ستجد رابط 'نسخة مخبأة' صغير في الأعلى، اختره لترى صور جوجل للصفحة. وإذا واجهت مشكلة في تحميلها، بإمكانك التحويل لترى الصفحة النصية الأكثر بدائية عن طريق اختيار رابط آخر في أعلى صفحة ذاكرة التخزين المؤقت الكاملة. سترغب في أخذ لقطة للشاشة أو نسخ ولصق أي محتوى ذي صلة، لأنه قد يتم إلغاؤه في أي وقت بواسطة الزحف التالي.

آلة الزمن لأرشيف الإنترنت Wayback Machine

إذا أردت أن تعرف كيف تغيرت صفحة معينة عبر فترة زمنية طويلة، كالأشهر أو السنوات، يقدم أرشيف الإنترنت خدمة تُدعى 'آلة الزمن' Wayback Machine (http://archive.org/web/web.php) والتي تأخذ لقطات لأكثر صفحات الويب رواجاً بشكل دوري. اذهب إلى الموقع وأدخل الرابط الذي ترغب في البحث عنه، وإذا كان لدى الأرشيف أي نسخ منه، سيظهر لك تقويم لتختار الزمن الذي تود معاينته، ثم سيقدم لك نسخة من الصفحة كما كانت تقريبًا في ذلك الوقت. غالبًا ما ستفتقد التصميم أو الصور، ولكنها تكفي لتفهم ما كان يركز عليه محتوى الصفحة حينذاك.

خيار View Source

هذا أمر قليل الاحتمالات، ولكن غالباً ما يترك المطوّرون تعليقات أو أدلة أخرى في رمز HTML الذي يكمن في أساس أي صفحة. ستجد الرمز في قوائم مختلفة حسب متصفحك، ولكن هناك دائماً خيار 'View Source' الذي يتيح لك تصفح HTML الخام. لستَ بحاجة إلى فهم ما تعنيه الأجزاء المقروءة آليًا، فقط راقب أجزاء النص التي عادة ما تكون متناثرة بينها. حتى وإن كانت مجرد إعلانات لحقوق النشر أو تنويهات عن أسماء المؤلف، فإنها غالبًا ما تعطيك أدلة هامة بشأن إنشاء الصفحة والغرض منها.

محرك بحث TinEye

أحيانًا ترغب في معرفة مصدر صورة ما، ولكن بدون أي نص يشير إلى مصدرها، لا توجد أي وسيلة واضحة للمعرفة باستخدام محركات البحث التقليدية مثل جوجل. يقدم محرك البحث TinEye (http://www.tineye.com/) عملية متخصصة لـ 'البحث العكسي عن الصور' حيث تعطيه الصورة ثم يعثر لك على صور شبيهة على الويب. ولأنهم يستخدمون خاصية التعرف على الصور ليقوموا بالمطابقة، ينجح الأمر حتى وإن تم قص الصورة أو تشويهها أو ضغطها. يمكن لهذا أن يكون فعالاً للغاية عندما تشك في تحريف صورة على أنها أصلية

أو جديدة، لأن الخدمة تعود بك إلى المصدر الحقيقي.

موقع YouTube

إذا اخترت أيقونة ’الإحصاءات‘ في أسفل اليمين لأي مقطع فيديو، سيكون بإمكانك الحصول على مجموعة غنية من المعلومات بشأن جمهور المقطع عبر الزمن. وعلى الرغم من أنها ليست كاملة، فإنها تفيد في فهم من هم المشاهدون وبلدانهم وزمن مشاهدتهم للمقطع.

البريد الإلكتروني

إذا كانت لديك بعض رسائل البريد الإلكتروني التي تبحث فيها، سترغب غالبًا في معرفة المزيد من التفاصيل عن هوية المرسل وموقعه. لا توجد أداة جيدة وجاهزة لتساعدك في ذلك، ولكن قد يكون من المفيد جدًا أن تعرف الأساسيات بشأن العناوين الخفية التي توجد في كل رسالة. تعمل تلك العناوين الخفية مثل ختم البريد، فيمكنها الكشف عن كم مفاجئ من تفاصيل المرسل. كما أنها تشمل على وجه الخصوص عنوان IP الخاص بالجهاز الذي أُرسلت منه الرسالة، فهي تشبه هوية المتصل في مكالمات الهاتف. يمكنك حينها أن تدخل عنوان IP هذا في موقع whois لتعرف أي منظمة تملك هذا الجهاز. وإذا تبين أن المالك Com-cast أو AT&T اللتان توفران خدمات الاتصال للمستهلكين، يمكن أن تزور موقع MaxMind لتعثر على المكان التقريبي. لترى تلك العناوين الخفية في Gmail، افتح الرسالة والقائمة التي توجد بجانب خيار الرد في أعلى اليمين واختر ‘إظهار الأصلية’. سترى صفحة جديدة تكشف عن المحتوى الخفي، وستكون هناك بضع عشرات من السطور في البداية مكونة من كلمات تتبعها نقطتان. قد يكون عنوان IP الذي تبحث عنه موجوداً في تلك السطور، ولكن سيعتمد الاسم على كيفية إرسال الرسالة. فإن كانت قد أرسلت من Hotmail، سيُدعى ‘X-Originating-IP:’ ولكن إن كانت من Outlook أو Yahoo، سيكون في السطر الأول الذي يبدأ بـ ‘Received:’.عند إدخال العنوان في موقع whois، يفيدنا بأنه يتبع Virgin Media، وهي شركة تزود خدمة الإنترنت في المملكة المتحدة، لذا عند إدخاله

في خدمة الموقع الجغرافي لدى موقع MaxMind أدرك أن مصدرها هو مدينة كامبريدج - مسقط رأسي - مما يجعلني أثق بقدر معقول أن والديَّ هما من أرسل الرسالة، وليس أشخاصاً محتالين!

الاتجاهات السائدة

إذا كنت تبحث في موضوع واسع وليس عن موقع أو أمر معين، ففيما يلي بعض الأدوات التي بإمكانها تزويدك ببعض الأفكار:

معدل الزيارات إلى مقالات ويكيبيديا

إذا كنت مهتمًا بمعرفة مدى تغيُّر الاهتمام العام بموضوع أو بشخص معين عبر الزمن، يمكنك أن تحصل على أرقام الزيارات اليومية لأي صفحة على ويكيبيديا من (http://stats.grok.se/). وهو موقع **غير معقد** ولكنه يفي بالغرض وسيتيح لك الكشف عما تحتاجه من معلومات بالقليل من التنقيب.

أدخل الاسم الذي يهمك لتحصل على عرض شهري لمعدل الزيارات لتلك الصفحة. سيظهر لك رسماً بيانياً يوضح عدد زيارات الصفحة لكل يوم من الشهر الذي حددته. للأسف يمكنك أن ترى فقط كل شهر على حدة، لذا ستضطر إلى اختيار شهر جديد وتبحث مجددًا لترى التغيرات طويلة الأمد.

خدمة Google Insights

بإمكانك الحصول على رؤية واضحة للعادات البحثية للناس عن طريق خدمة Insights من جوجل (http://www.google.com/insights/search/) (نموذج 11-4). أدخل بعض الجُمل البحثية الشائعة مثل 'جاستن بيبر أَم ليدي جاجا' وسترى رسمًا بيانيًا يوضح الأرقام النسبية لعدد مرات البحث عنها عبر الزمن. هناك العديد من الخيارات لتنقية رؤيتك للبيانات، بدءًا من المناطق الجغرافية ذات النطاق الأضيق وصولًا إلى المزيد من التفاصيل عبر الزمن. الأمر الوحيد الذي يخيب الآمال في تلك الخدمة هو عدم وجود قيم مطلقة، فأنت تحصل فقط على أرقام نسبية يصعب تفسيرها.



نموذج 4–11. خدمة Google Insights (جوجل)

— بيت واردن، محلل ومطوّر بيانات مستقل

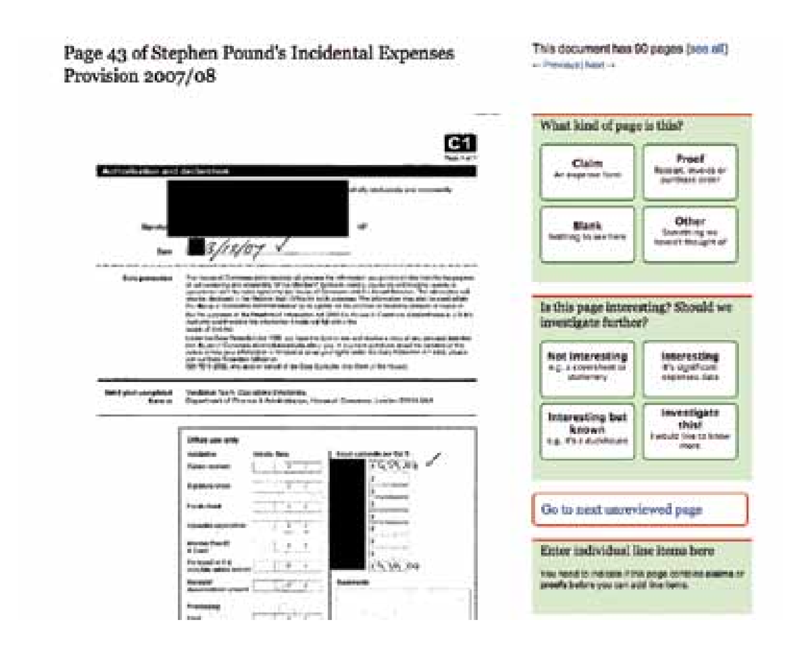
التعهيد الجماعي للبيانات في مدونة بيانات صحيفة الغارديان

التعهيد (التكليف) الجماعي - وفقاً لويكيبيديا الإنجليزية (http://en.wikipedia.org/wiki/Crowdsourcing) - هو "عملية موزعة لحل المشاكل والإنتاج وتشمل إسناد المهام إلى شبكة ممن يُطلق عليهم أيضًا اسم 'الجماهير'". وفيما يلي جزء من مقابلة مع سايمن روجرز يوضح فيها كيف استخدمت مدونة البيانات التعهيد الجماعي لتغطية فضائح نفقات النواب وتعاطي المخدرات وأوراق سارة بالين:

في بعض الأحيان، قد تحصل على أطنان من الملفات أو الإحصاءات أو التقارير التي يستحيل على شخص واحد أن يتفحصها. وقد تحصل أيضًا على مواد لا يمكن النفاذ إليها أو في صيغة سيئة ولن تتمكن من الاستفادة منها. وهنا يمكن للتعهيد الجماعي أن يساعدك.

لدى صحيفة الغارديان الكثير من القراء والكثير من العيون. فإذا كان هناك مشروع مثير للاهتمام نحتاج فيه إلى بعض المُدخلات، يمكننا أن نطلب منهم المساعدة. وهذا ما فعلناه في قضية نفقات النواب (http://mps-expenses.guardian.co.uk/) حيث كانت لدينا 450,000 وثيقة وكان وقتنا محدودًا للغاية

للقيام بأي شيء. وبالتالي لم يكن هناك أفضل من إشراك قرائنا للمساهمة في إنجاز المهمة.



نموذج 4–12. نسخة منقحة من النفقات الطارئة لستيفن باوند (الغارديان)

نجم عن مشروع نفقات النواب الحصول على الكثير من البلاغات التي تضمنت معلومات سرية، وتلقينا العديد من الأخبار أكثر من البيانات. وشهد ذلك المشروع نجاحًا ملحوظًا فيما يتعلق بمعدل الزيارات، فلقد أعجب الناس أيما إعجاب.

نعمل الآن بالتعاون مع MixMag على مشروع حول تعاطي المخدرات (http://bit.ly/guardian-drugs)، وهو موضوع متميز كذلك. ويبدو أنه سيكون أكبر من الدراسة المسحية للجريمة في بريطانيا، بسبب عدد الذين عادوا لتعاطيها، وهذا أمر مثير.

يشترك كلا المشروعين في كونهما يتعلقان بقضايا يهتم بها الناس فعلاً، لذا فلديهم الاستعداد لقضاء الوقت في العمل عليها. وقد اعتمد الكثير من التعهيد الجماعي الذي قمنا به على المساعدة من الأشخاص المهووسين بالموضوع. ففي

مشروع نفقات النواب، حصلنا على معدل هائل من الزيارات في البداية ثم قلّت تدريجيًا. ولكن لا يزال لدينا بعض المهووسين بتصفح كل صفحة بحثاً عن الأخطاء والأخبار. وقد قام أحد الأشخاص بتصفح 30,000 صفحة. لذا فهم يعرفون الكثير.

استخدمنا التعهيد الجماعي أيضاً في مشروع أوراق سارة بالين (http://bit.ly/guardian-palin-papers). ومرة أخرى ساعدنا هذا كثيراً في البحث عن الأخبار من خلال المعلومات.

وقد كان التعهيد الجماعي ذا نفع كبير لنا فيما يخص الأخبار. فقد أعجَب الناس كثيرًا مما حسّن مظهر الغارديان. ولكن فيما يخص البيانات، لم نستخدم التعهيد الجماعي كثيراً.

بعض مشاريع التعهيد الجماعي التي قمنا بها وحققت نجاحًا كانت أشبه بدراسات تقليدية. فعندما تسأل الناس عن تجاربهم وحياتهم وما فعلوه، يعملون بشكل جيد جدًا لأنهم لن يختلقوا ذلك بل سيقولون ما يشعرون به. وعندما طلبنا من الناس أن يقوموا بجزءٍ من عملنا، كان علينا أن نجد إطارًا لهؤلاء الناس يعملون من خلاله لإنتاج بيانات يمكن الوثوق بها.

أما بالنسبة لموثوقية البيانات، فأرى أن النهج الذي يتبعه مشروع Old Weather (http://www.oldweather.org/) جيد للغاية. فهُم يكلفون عشرة أشخاص بالعمل على كل مُدخل، وهذه طريقة جيدة لضمان الدقة. وفي مشروع نفقات النواب، حاولنا تقليل خطر دخول النواب إلى الموقع وتعديل سجلاتهم الخاصة لتحسين صورتهم. ولكن لا يمكنك مكافحة ذلك دائما، يمكنك فقط أن تتوخى الحذر من عناوين URL معينة أو إذا كان مصدرها المنطقة ذات الرمز البريدي SW1 (مناطق تمركز أصحاب النفوذ السياسي والمالي) في لندن. لذا فالأمر أكثر صعوبة، فالبيانات التي كنا نستخرجها لم تكن موثوقة دائمًا. وعلى الرغم من أن الأخبار كانت متميزة، إلا أنها لم تقدم أرقاماً موثوقة يمكننا استخدامها.

وإن طلب مني تقديم النصح لصحافيي البيانات الطامحين في استخدام التعهيد الجماعي لجمع البيانات، سأشجعهم على فعل ذلك على أن تكون القضية محط اهتمام بالغ من الجمهور، حيث سيواصلون الاهتمام بها حتى بعد أن تتوقف عن تصدّر العناوين الرئيسية. وإذا قمت بشيء مثل لعبة، على سبيل المثال، فإن ذلك

سوف يساعد في إشراك عدد كبير من الناس. فعندما عملنا على خبر النفقات مرة أخرى، كان الأمر أشبه بلعبة مع مهام فردية يؤديها الناس. وقد ساعدنا هذا كثيرًا في تكليف الناس بمهام معينة، مما أحدث أثراً إيجابياً كبيراً. فأنا أعتقد أنك إذا لم تقدم للناس سوى تلٍّ من المعلومات ثم طلبت منهم أن يتفحصوه، سيرونه عملا شاقا ومملا وغير مجز. لذا أعتقد أنه من المهم جعل المهمة ممتعة ومثيرة لهم.

—— ماريان بوشار، مدونة صحافة البيانات، في مقابلة مع سايمون روجرز، صحيفة الغارديان

كيف استخدمت مدونة البيانات التعهيد الجماعي لتغطية تذاكر الأولمبياد

أظن أن مشروع التعهيد الجماعي الذي حصل على أكبر استجابة كان عن اقتراع تذاكر أولمبياد لندن (http://bit.ly/guardian–olympics). فقد حاول الآلاف في المملكة المتحدة الحصول على تذاكر أولمبياد 2012 وكان هناك غضب عارم بشأن عدم حصول الناس عليها. وقد طلب الناس تذاكر بمئات الجنيهات وقيل لهم أنهم لن يحصلوا على أي شيء. ولكن لم يعرف أحد ما إذا كان كل ذلك مجرد شكوى صاخبة من البعض بينما كان معظم الناس سعداء. لذا حاولنا إيجاد طريقة لمعرفة ذلك.

قررنا أن أفضل ما يمكننا فعله في ضوء افتقاد أي بيانات جيدة حول الموضوع هو سؤال الناس. وقد اعتقدنا أننا سنضطر إلى التعامل مع الأمر على أنه موضوع اجتماعي خفيف، لأن العينة لم تكن متوازنة.

قمنا بابتكار استمارة جوجل وطرحنا أسئلة محددة جدًا (http://bit.ly/guardian–olympics2). كانت الاستمارة طويلة وتضمنت أسئلة حول قيمة ما طلبه الناس من تذاكر وما قيمة المبلغ المخصوم من بطاقاتهم وأي الفعاليات اختاروا وما إلى ذلك.

وضعنا الاستمارة كصورة صغيرة في واجهة الموقع وتم تعميمها بسرعة. أعتقد أن هذه كانت من أهم الخطوات، فلا يمكن أن تفكر فقط فيما 'تود معرفته من أجل الخبر'، بل عليك أن تفكر فيما 'يود الناس إخبارك به الآن'. وفقط عندما

How many Olympic tickets did you get? Here's our readers' results

We asked how you had fared in the London 2012 ticket ballot. Here's our analysis of the information you gave us





The London 2012 Olympic stadium: will you be there? Will anyone you know? Photograph: Handout/Getty Images



Posted by
James Ball
Friday 3 June 2011 11.04 BST
guardian.co.uk
Article history






Sport
Olympic tickets · Olympic Games 2012

More from London 2012 Olympics blog on

Sport
Olympic tickets · Olympic Games 2012

More blogposts

How much money did London 2012 ticket buyers have to put on the line to stand even a 50:50 chance of getting at least one ticket? If the results submitted by Guardian readers are to be believed, at least £1,000.

Earlier this week, we asked readers of the Guardian London 2012 blog to let us know how their ticket-purchasing attempts had fared. By the end of the day – when this analysis was carried out – we'd had more than 5,000 responses.

نموذج 4–13. كم تذكرة أولمبياد حصلت عليها؟ نتائج القراء (الغارديان)

تتطرق إلى ما يود الناس الحديث عنه، ينجح مشروع التعهيد الجماعي. تلقى هذا المشروع – إحدى أولى تجاربنا في التعهيد الجماعي – كمًا هائلًا من الردود، فتلقينا ألف رد في أقل من ساعة ومع نهاية اليوم بلغت الردود سبعة آلاف.

لذا بالطبع أخذنا تقديم النتائج على محمل الجد في تلك المرحلة. في البدء لم تكن لدينا أدنى فكرة عن مدى نجاح المشروع. لذا أضفنا بعض التنويهات: أن قراء الغارديان قد يكونون أكثر ثراء من الآخرين، وأن هؤلاء الذين حصلوا على أقل مما توقعوه قد يكون لديهم استعداد أكبر للتحدث إلينا، وهكذا.

لم نكن نعرف القيمة التي سترِد في النتائج، فانتهى بنا الأمر إلى الحصول على سبعة آلاف نتيجة لنبني عليها مقالنا، واكتشفنا أن نصف الذين طلبوا التذاكر

لم يتلقوا شيئًا. درسنا كل تلك الأمور وكان هناك اهتمام كبير بالنتائج لأن العديد من الناس قد شاركوا في العمل في اليوم السابق.

بعد بضعة أسابيع، صدر التقرير الموجز الرسمي وقد كانت أرقامنا متقاربة بشكل صادم، فقد كانت مضبوطة تقريبًا. أعتقد أن جزءًا من السبب كان حسن الحظ ولكن أيضًا لأننا تعاملنا مع الكثير من الناس.

إذا بدأت تسأل قراءك عن أمر كهذا في سلسلة من التعليقات، سيكون ما تفعله بالنتائج محدودًا. لذا عليك أن تبدأ بالتفكير: 'ما هي أفضل أداة لما أريد معرفته؟' هل هي سلسلة التعليقات؟ أم هي بناء تطبيق خاص؟ وإذا كان المطلوب بناء التطبيق، عليك أن تفكر: 'هل الأمر يستحق الانتظار؟ وهل يستحق الموارد المطلوبة لتنفيذه؟'

في تلك الحالة فكرنا في استمارات جوجل، لأنه إذا ملأها أحدهم سترى النتائج مصفوفة في جدول بياني. وهذا يعني أنه حتى وإن كانت الأجوبة لا تزال في طور التحديث، وحتى وإن كانت النتائج لا تزال تتوالى، يمكنني أن أفتح الجدول البياني وأرى النتائج فورًا.

كان بإمكاني محاولة القيام بالعمل داخل تطبيق جوجل ولكنني قمت بتحميل البيانات في صيغة ملف إكسل ثم رتبته من أسفل لأعلى ووجدت من قرروا الكتابة بدلاً من وضع الأرقام بشأن ما أنفقوه فقمت بمعالجة الأمر. قررت بعد ذلك أن أتفادى الإقصاء قدر المستطاع، لذا بدلاً من أخذ الإجابات الصالحة فقط، حاولت معالجة الإجابات الأخرى. وقد استخدم بعض الناس عملات أجنبية فقمت بتحويلها إلى الجنيه الاسترليني، وكانت تلك مهمة شاقة.

ولكن التحليل بأكمله تم إنجازه في بضع ساعات بعد أن تخلصت من الإجابات السخيفة. فقد قرر الكثيرون أن يكتبوا أنهم لم ينفقوا شيئًا على التذاكر، وهذا طريف بعض الشيء ولكن لا بأس به، فقد شكلت تلك الإجابات أقل من مائة إجابة من بين سبعة آلاف مشارك.

كان هناك بضع عشرات ممن وضعوا أرقامًا مرتفعة– عشرة ملايين جنيه مثلاً – وكان من الواضح أنها زائفة في محاولة منهم لتشويه النتائج. فبعد استبعاد كل ذلك، تبقت مجموعة يمكنني استخدامها مع مبادئ البيانات العادية التي نستخدمها

يوميًا. وضعت ما يسمى بـ ’جدول محوري‘ ووضعت بعض النسب المتوسطة وما إلى ذلك من أمور.

لم تكن لدينا أدنى فكرة عن مدى الزخم الذي سيحدثه المشروع، لذا فقد عملت مع محرر مدونة الرياضة فقط، حيث تعاونّا معًا وتوقعنا أن يكون مشروعا ممتعًا. وقد نجحنا في إنجاز المشروع من الألف إلى الياء في 24 ساعة. كانت لدينا الفكرة، فوضعنا شيئًا في الصفحة الرئيسية للموقع ظهرًا. ووجدنا أن المشروع قد أصبح رائجًا، فأبقيناه في الصفحة الرئيسية لبقية اليوم ونشرنا النتائج على الإنترنت في صباح اليوم التالي.

قررنا استخدام محرر مستندات جوجل لأنه يتيح لنا التحكم الكامل في النتائج. لم أضطر إلى استخدام أي الأدوات التحليلية الأخرى حيث استطعت إدخاله بسهولة إلى برنامج قاعدة بيانات أو جداول بيانية. وعندما تبدأ في استخدام برامج التصويت المتخصصة، ستتقيد غالبًا باستخدام أدواتها. ولو كانت المعلومات التي نطلبها حساسة لترددنا قبل استخدام جوجل وفكرنا في العمل ’داخليًا‘. ولكن عمومًا، من السهل وضع استمارة جوجل في صفحة الغارديان، وبالنسبة للمستخدم فمن غير المرئي تقريبًا أننا نستخدمها، مما يجعلها مناسبة جدًا لهذا الغرض.

أما بالنسبة للنصائح لصحافيي البيانات الذين يرغبون في استخدام التعهيد الجماعي فعليك أن تحدد بالضبط ما تريد معرفته. اطرح أسئلة قد تحصل على إجابات متعددة الخيارات قدر المستطاع. حاول أن تعرف التركيبة الديموغرافية الأساسية لمن تخاطبهم حتى تتبين إن كانت عينتك متحيزة. وإذا كنت تطلب معرفة كميات وما إلى ذلك من أمور، حاول أن تحدد في الإرشادات أن تكون في صيغة أرقام أو أن عليهم استخدام عملة معينة وهكذا. فكلما ساعدتهم، كان ذلك أفضل. واحرص دائمًا على إضافة خانة للتعليقات لأن الكثيرين سيملأون الخانات الأخرى ولكن ما يريدونه حقاً هو الإدلاء بآرائهم عن الخبر. وخاصة إذا كان الخبر يخص المستهلكين أو الغضب العام.

— ماريان بوشار، مدونة صحافة البيانات، في مقابلة مع جيمس بال، صحيفة الغارديان

استخدام البيانات وتعميمها: النص الأساسي والملاحظات الثانوية والواقع

في هذا الباب سنلقي نظرة سريعة على القانون فيما يتعلق بالبيانات وقواعد البيانات، وما يمكنك فعله للوصول إلى بيانات تريدها باستخدام تراخيص وأدوات قانونية جاهزة ومتاحة للعامة. لا تسمح لذلك بتثبيط حماسك الخاص بالصحافة المبنية على البيانات. فالقيود القانونية على البيانات عادة لن تقف في طريقك، ويمكنك بسهولة أن تحرص على عدم وقوفها في طريق الآخرين ممن يستخدمون البيانات التي تنشرها.

من نافلة القول، أن الحصول على البيانات لم يكن متوفرًا في أي وقت بتلك السهولة كما هو الآن. فقبل النشر الواسع للبيانات على الإنترنت، حتى وإن حددت مجموعة البيانات التي تحتاج إليها، ستحتاج إلى أن تطلب ممن يملك نسخة منها أن يتيحها لك، وقد يتطلب ذلك تحضير الأوراق أو استخدام البريد أو القيام بزيارة شخصية. والآن اجعل جهازك يطلب من جهازهم إرسال نسخة إلى جهازك! وهذا أمر مشابه نظريًا، ولكن الآن لديك نسخة، بينما هُم (صاحب البيانات أو ناشرها) لم يفعلوا شيئًا، وغالبًا ليست لديهم أدنى فكرة عن قيامك بتحميل النسخة.

ماذا عن تحميل البيانات باستخدام برنامج (يُسمى أحيانًا 'أداة استخراج البيانات من الويب' 'scraper') وشروط الخدمة ToS؟ فكر في الفقرة السابقة. إذا كان لديك كم من الوقت والمال لقراءة مثل تلك الوثائق أو ربما طلب النصيحة من محام ما، إذن فافعل ذلك، ولكن لا تكن أحمق: لأنه إذا اقتحم برنامجك أحد المواقع، فقد يتم منع شبكتك من الدخول إلى الموقع المذكور وربما تستحق ذلك. هناك الآن مجموعة من الممارسات الخاصة بالحصول على البيانات واستخراجها من الويب. فإذا كنت تنوي فعل ذلك، ستعطيك قراءة الأمثلة على موقع مثل ScraperWiki انطلاقة قوية.

بمجرد حصولك على بعض البيانات التي تهمك، يمكنك تفحصها والتدقيق فيها وفرزها وتوضيحها في صورة عروض مرئية وربطها والقيام بأي نوع آخر من التحليل الذي ترغب فيه مستخدمًا نسختك من البيانات. وتستطيع نشر تحليلك

واقتباس أي من المعطيات. هناك الكثير وراء شعار ’حرية الحقائق‘ (كما في حرية التعبير)، ولكنه قد يكون مجرد شعار بالنسبة للذين يفكرون كثيرًا في الجوانب القانونية لقواعد البيانات، أو حتى على نحو أشمل (أو غير دقيق)، لإدارة البيانات.

ماذا لو كنت تنوي – بصفتك صحفي بيانات جيدًا أو تطمح لأن تكون جيدًا – ألا تنشر تحليلك بما في ذلك بعض الحقائق أو نقاط البيانات فحسب، بل ستنشر كذلك مجموعات وقواعد البيانات التي استخدمتها – وربما أضفت إليها – عند إجراء تحليلك؟ أو ربما تجمع البيانات فقط ولم تجرِ أي تحليلات وهذا أمر جيد لأن العالم يحتاج إلى من يجمع البيانات لحمايتها. وإذا كنت تستخدم بيانات جمعتها جهات أخرى، قد تواجه عقبة. (إذا كنت أنت من جمع قاعدة بياناتك بأكملها، اقرأ الفقرة التالية عموماً كحافز لممارسات التعميم في الفقرة التي تليها).

إذا كان لديك علم عن كيفية تقييد حقوق النشر للأعمال الإبداعية، وإذا لم يسمح صاحب حقوق النشر باستخدام أحد الأعمال (أو إذا كان العمل في المجال العام أو إن كانت ستُطبق على استخدامك بعض الاستثناءات والقيود مثل سياسة الاستخدام العادل) وقُمت باستخدام – أو توزيع أو تأدية، إلخ – العمل بغض النظر عن كل ذلك، يستطيع صاحب حقوق النشر أن يجبرك على التوقف. وعلى الرغم من عدم وجود قيود على نشر الحقائق، يمكن تقييد جمع الحقائق على نحو مماثل، رغم أن هناك تباينًا في القوانين ذات الصلة بالحقائق أكثر منه في تطبيق حقوق النشر على الأعمال الإبداعية. باختصار، يمكن تطبيق حقوق النشر على قاعدة البيانات بصفتها عملًا إبداعيًا. بموجب قانون ’عرق الجبين‘ ترى العديد من الولايات القضائية أن مجرد تجميع قاعدة بيانات، حتى وإن لم تكن بشكل إبداعي، يجعلها مرهونة بحقوق النشر. وفي الولايات المتحدة على وجه الخصوص، يتطلب تطبيق حقوق النشر مستوى أعلى من الحد الأدنى للإبداع (قضية شركة فيست Feist التي تتعلق بدليل الهاتف، وهي قضية نموذجية في الولايات المتحدة إذا أردت أن تبحث عنها). ولكن في بعض الولايات القضائية، هناك أيضًا ’حقوق قواعد البيانات‘ التي تضع قيودًا على قواعد البيانات بشكل منفصل عن حقوق النشر (رغم أن هناك الكثير من التداخل فيما يتعلق بما تتم تغطيته، وخاصة عندما ينعدم تقريبًا الحد الأدنى للإبداع في حقوق النشر). وأفضل مثال معروف على ذلك

هو حقوق قواعد البيانات ’الفريدة من نوعها‘ لدى الاتحاد الأوروبي. أعود وأكرر، إن كنت في أوروبا على وجه الخصوص، ينبغي أن تتأكد من حصولك على الإذن قبل نشر قاعدة بيانات مأخوذة من أي كيان آخر.

بالطبع ليست هذه القيود أفضل وسيلة لبناء نظام متكامل لصحافة البيانات (وليست جيدة بالنسبة للمجتمع ككل، فقد قال الباحثون الاجتماعيون وآخرون للاتحاد الأوروبي أن تلك القيود ليست جيدة قبل صدور قاعدة البيانات ’الفريدة من نوعها‘، ومنذ ذلك الحين أظهرت الدراسات أنهم محقون). لحسن الحظ، عندما تكون ناشر قاعدة البيانات، يمكنك إزالة مثل تلك القيود منها (على افتراض أنها لا تحتوي على عناصر لا يُسمح لك بنشرها) من خلال منح الإذن مقدمًا. تستطيع القيام بذلك عن طريق نشر قاعدة البيانات بترخيص عام أو بإهدائها للجمهور مثلما ينشر الكثير من المبرمجين الترميز (الكود) الخاص بهم بترخيص مجاني ومفتوح المصدر، وذلك حتى يتمكن الآخرون من الإضافة إلى الكود الخاص بهم (لأن صحافة البيانات غالبًا ما تشمل الرمز وليس البيانات فقط، وبالطبع عليك أن تنشر الرمز كذلك ليكون جمعك للبيانات وتحليلها قابلين للنسخ). هناك الكثير من الأسباب لنشر بياناتك. فمثلاً، قد يصنع جمهورك توضيحات مرئية جديدة أو تطبيقات يمكن الربط بها – وهذا ما تفعله صحيفة الغارديان في مجموعة التوضيحات المرئية الخاصة بها على موقع فليكر. يمكن إضافة مجموعات بياناتك إلى مجموعات بيانات أخرى لترى أنت وقراؤك المزيد من الأفكار حول الموضوع. وما يفعله الآخرون ببيانات قد يزودك بخيوط تؤدي إلى أخبار جديدة أو أفكار لأخبار أو مشاريع أخرى تعتمد على البيانات. كما سيجلبون لك الشهرة بالتأكيد.

عندما يدرك المرء أن نشر الأعمال بتراخيص عامة أمر ضروري، يصبح السؤال: أي ترخيص؟ هذا السؤال الصعب ستتم الإجابة عليه كثيرًا من قِبَل المشروع أو الجماعة التي تضيف عملك إلى عملهم، أو الذين تأمل في إضافة عملك إليهم، وهي: استخدم الترخيص الذي يستخدمونه. إذا احتجت إلى التعمق في البحث، أبدأ بمجموعة التراخيص المجانية والمفتوحة – أي أن الجميع لديهم الإذن، أيًا كان استخدامهم (قد يتطلب الأمر أن ينسب العمل إلى صاحبه والمشاركة بالمثل). وما تقوم به وثيقتا ’تعريف البرامج المجانية‘ و’تعريف المصدر المفتوح‘



نموذج 4–14. بطاقات البيانات المفتوحة (مؤسسة المعرفة المفتوحة)

للبرامج، تقوم به وثيقة ’تعريف المعرفة المفتوحة‘ (http://opendefinition.org/) لكل أشكال المعرفة الأخرى، بما فيها قواعد البيانات: حيث يحدد ما يجعل العمل مفتوحًا، وما تسمح به التراخيص المفتوحة للمستخدمين.

يمكنك زيارة موقع ’تعريف المعرفة المفتوحة‘ لترى مجموعة التراخيص المؤهلة حاليًا (http://opendefinition.org/licenses/). وباختصار، توجد ثلاثة تصنيفات أساسية للتراخيص المفتوحة:

- الإهداء إلى الجمهور، وهو ترخيص مجيز إلى أقصى حد، ولا توجد شروط لاستخدام العمل.
- التراخيص المجيزة أو التي تتطلب النسبة فقط، حيث تكون نسبة العمل إلى صاحبه الشرط الأساسي الوحيد.
- ترخيص الحقوق المتروكة والتراخيص المتبادلة والمشاركة بالمثل: تتطلب تلك أيضًا مشاركة الأعمال التي تم تعديلها – إذا نُشرت – بنفس الترخيص.

لاحظ أنك إذا كنت تستخدم قاعدة بيانات نشرها شخص آخر بترخيص مفتوح، اعتبر الفقرة السابقة دليلاً موجزًا لكيفية الوفاء بشروط هذا الترخيص المفتوح. عادة ما تحتوي التراخيص التي من الأرجح أن تجدها من منظمة المشاع الإبداعي Creative Commons ومشاع البيانات المفتوحة على ملخص يسمح لك بأن ترى الشروط الأساسية بسهولة. وعادة ستجد الترخيص على صفحة الويب التي يتم تحميل مجموعة البيانات منها (أو 'استخراجها' لأن صفحات الويب يمكن بالطبع أن تحتوي على مجموعات البيانات) أو في مكان واضح داخل مجموعة البيانات نفسها حسب الصيغة. وهذا ما يجب أن تفعله أيضًا عندما تتيح مجموعات بياناتك.

عَوْدٌ على بَدْء، ماذا لو كانت مجموعة البيانات التي تحتاج إليها لا تزال غير متاحة على الإنترنت، أو لو كانت تخضع لبعض أنواع التحكم في الوصول؟ إضافة إلى طلب الوصول إليها بنفسك، ليكن في اعتبارك طلب إتاحتها للعالم بأسره. فعليك أن تلفت انتباههم إلى بعض الآثار الجيدة المتوقعة عندما يسمحون بإتاحة بياناتهم.

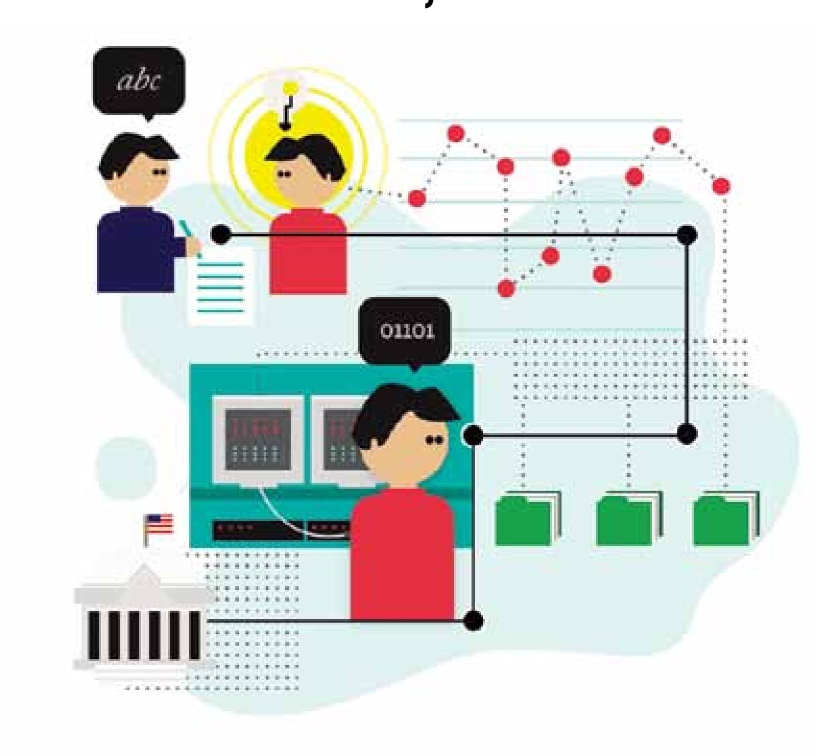
قد يرد بخواطرنا أن تعميم البيانات حول العالم سيجعل من الخصوصية والاعتبارات والقواعد الأخرى في بعض مجموعات البيانات مادة للعب. ولكن نؤكد أن خفض البيانات المفتوحة لسقف العديد من الحواجز التقنية والتي تخص حقوق النشر وما يشبهها، لا يعني أنه ليس عليك أن تخضع للقوانين المطبقة الأخرى. ولكن كثيراً ما كان الوضع هكذا، كما أن هناك كمًا هائلًا من الموارد وفي بعض الأحيان أنواع من الحماية للصحافيين إذا شعرت بضرورة التحقق من ذلك.

أتمنى لك حظًا سعيدًا! ولكن على الأرجح سوف تحتاج إلى الكثير من الحظ في أجزاء أخرى من مشروعك أكثر منه في إدارة المخاطر القانونية (المنخفضة).

— مايك لينكسفاير، منظمة المشاع الإبداعي

الفَصْل الخَامِس

فهم البيانات



ماذا تفعل بالبيانات بعد الحصول عليها؟ ما الذي يجب أن تبحث عنه؟ ما هي الأدوات التي يجب أن تستخدمها؟ يبدأ هذا الباب ببعض الأفكار حول تحسين إلمامك بالبيانات ونصائح للعمل مع الأرقام والإحصاءات وأمور يجب وضعها في الاعتبار عندما تعمل مع مجموعات البيانات الفوضوية والناقصة والتي عادة ما تكون غير موثّقة. نواصل التعرف على كيفية الحصول على الأخبار من البيانات والأدوات التي يختارها صحافيو البيانات وكيفية استخدام التوضيح المرئي لتزويدك بأفكار حول الموضوع الذي تبحث فيه.

تعلّم البيانات في ثلاث خطوات بسيطة

مثلما يشير التعلُّم إلى "اكتساب القدرة على القراءة من أجل المعرفة، والكتابة بشكل متسق، والتفكير الناقد" في المواد المطبوعة، فإن تعلُّم البيانات يعني القدرة على استخدامها من أجل المعرفة والإنتاج بشكل متسق والتفكير بمنظور نقدي في البيانات. ويشمل تعلُّم البيانات الإلمام بالإحصاءات، وفهم كيفية التعامل مع مجموعات بيانات ضخمة، وكيف يتم إنتاجها، وكيفية الربط بين مجموعات بيانات عديدة وتفسيرها.



نموذج 5–1. التنقيب في البيانات (تصوير: جي دي هانكوك J. D. Hancock، http://www. flickr.com/photos/jdhancock/3386035827/)

تقدم جامعة نيوز التابعة لمعهد بوينتر الإعلامي دروسًا في الرياضيات للصحافيين (http://www.newsu.org/courses/math-journalists)، حيث يحصل الصحافيون على المساعدة في تعلم مفاهيم رياضية مثل تغيُّر النسب المئوية والمتوسط الحسابي. ومما يثير الاهتمام أن هذه المفاهيم تُدرَّس في الوقت ذاته بالقرب من مكاتب بوينتر في مدارس فلوريدا لطُلاب الصف الخامس (الذين تتراوح أعمارهم بين 10- 11)، كما هو موضح في خطة المنهج الدراسي (http://bit.ly/k12-courses).

احتياج الصحافيين إلى المساعدة في دروس الرياضيات، التي عادة ما تُدرّس قبل المرحلة الثانوية، يظهر مدى بُعد غرف الأخبار عن تعلُّم البيانات، وهو أمر لا يخلو من المشاكل. إذ كيف يمكن لصحافي البيانات أن يستفيد من مجموعة أرقام عن التغيُر المناخي إذا لم يعرف معنى فترة الثقة؟ وكيف يمكنه كتابة خبر عن توزيع الدخل إذا لم يستطع معرفة الفرق بين الوسيط والمتوسط (http://bit.ly/karenberger-mean-median)؟

بالتأكيد لا يحتاج الصحفي إلى شهادة في الإحصاء ليصبح أكثر كفاءة في تعامله مع البيانات. وعندما يتعلق الأمر بالأرقام، توجد بعض الحيل اليسيرة التي تساعد في الحصول على خبر أفضل كثيرًا. وكما يقول غريد غيغرنزر، الأستاذ بمعهد ماكس بلانك، الحصول على أدوات أفضل لن يؤدي إلى صحافة أفضل إن استخدمت بدون رؤية.

حتى وإن كنت تفتقر إلى أي معرفة بالرياضيات أو الإحصاء، يمكنك بكل سهولة أن تصبح صحافي بيانات متمرساً من خلال طرح ثلاثة أسئلة بسيطة جدًا.

1- كيف جُمعت البيانات؟

نمو مدهش لإجمالي الناتج المحلي

أسهل الطرق للتباهي ببيانات مذهلة هو التزييف. يبدو هذا واضحًا، ولكن البيانات، التي تحظى بكثرة التعليقات من قبل العامة مثل أرقام الناتج المحلي الإجمالي، يمكن تزويرها بصورة احترافية. يقول السفير البريطاني السابق كريغ

موراي في كتابه 'جريمة قتل في سمرقند' (http://amzn.to/murder–samarkand)
إن معدلات النمو في أوزبكستان تخضع لمفاوضات مكثفة بين الحكومة المحلية والهيئات الدولية. وبعبارة أخرى، لا يتعلق الأمر بالاقتصاد المحلي.

يُستخدم الناتج المحلي الإجمالي كمؤشر أول لأن الحكومات تحتاج إليه لمراقبة مصدر الدخل الأساسي، ألا وهو ضريبة القيمة المضافة. وحين لا تعتمد الحكومة على هذه الضريبة في تمويلها أو حين لا تعلن الحكومة موازنتها أمام الجمهور، فلن يكون هناك سبب لدى الحكومات لجمع بيانات إجمالي الناتج المحلي، وسوف تكون أفضل حالاً بتزييف تلك البيانات.

معدلات الجريمة في ارتفاع متزايد

وفقاً لما جاء في صحيفة إل باييس (http://bit.ly/elpais–numeracy)، "ارتفع معدل الجريمة في إسبانيا بنسبة 3٪"، وحسب ما أفادته شبكة RTL (http:// bit.ly/rtl–numeracy)، سقطت مدينة بروكسل ضحية للجرائم المتزايدة بسبب المهاجرين غير الشرعيين ومدمني المخدرات. هذا النوع من التقارير المبنية على الإحصاءات التي جمعتها الشرطة أمر شائع، ولكنها لا تخبرنا الكثير عن العنف.

نثق بأن البيانات داخل الاتحاد الأوروبي لا يتم التلاعب بها، إلا أن أفراد الشرطة يستجيبون للحوافز. فعندما يتم ربط أداء رجال الشرطة بمعدلات الإدانة مثلاً، فإن ذلك يشكّل حافزًا لهم ليسجلوا، بقدر المستطاع، الحوادث التي لا تحتاج إلى التحقيقات، مثل تدخين الماريجوانا. وهذا يفسر سبب زيادة معدل الجرائم المتعلقة بالمخدرات في فرنسا إلى أربعة أضعاف خلال الخمس عشرة سنة الماضية بينما يظل معدل التعاطي ثابتًا.

ما يمكنك فعله

حينما لا تكون متأكدًا من مصداقية رقم ما، أعد التحقق منه دائمًا، كما كنت ستفعل لو ذكر الرقم في اقتباس من أحد السياسيين. بالنسبة للحالة الأوزبكية، يكفيك إجراء مكالمة هاتفية لشخص عاش هناك فترة لتتأكد من ذلك ("هل يبدو أن البلاد صارت أغنى ثلاثة أضعاف مقارنة بما كانت عليه عام 1995 كما تظهر الأرقام الرسمية؟).

أما بالنسبة لبيانات الشرطة، فعادة ما يقوم علماء الاجتماع بإجراء دراسات عن

ضحايا الجرائم حيث يسألون الناس إذا ما كانوا عرضة للجرائم. وهذه الدراسات أقل تقلبًا من بيانات الشرطة. وربما لذلك السبب لا تتصدر عناوين الأخبار.

هناك اختبارات أخرى تتيح لك تقييم مصداقية البيانات بدقة، مثل قانون بينفورد، ولكن لا يمكن لأي منها أن يحل محلَ تفكيرك النقدي.

2- ما الذي يمكن أن نتعلمه؟

يتضاعف خطر الإصابة بمرض التصلب المتعدد عند العمل ليلاً

بعد قراءة عنوان هذا الخبر (http://bit.ly/dmsg-numeracy)، بالتأكيد سوف يتوقف أي ألماني عاقل عن العمل في النوبات الليلية. ولكن المقال في النهاية لا يطلعنا على الخطر الحقيقي.

فمثلاً من بين كل ألف ألماني، سيصاب أحدهم بالمرض أثناء حياته. أما إذا عمل كل شخص من هؤلاء الألف في نوبات ليلية، فسيقفز عدد المصابين بالمرض إلى اثنين. إن الخطر الإضافي للإصابة بالمرض عند العمل في تلك النوبات هو واحد في الألف وليس 100٪. وبالتأكيد تزداد فائدة هذه المعلومات عندما تفكر فيما إذا كنت ستقبل بالوظيفة أم لا.

في المتوسط، 1 من بين كل 15 أوروبي أمّيٌّ تمامًا

يبدو هذا العنوان مثيرًا للذعر. وهذا صحيح تمامًا، فمن بين الخمسمئة مليون أوروبي، هناك 36 مليون شخص لا يمكنهم القراءة على الأرجح. ولكن هناك أيضاً 36 مليون تحت سن السابعة (البيانات من يوروستات Eurostat، http://bit.ly/eurostat-numeracy).

عندما تكتب عن المعدل المتوسط، فكّر دائمًا "متوسط ماذا؟" هل السكان الذين تشير إليهم متجانسون؟ على سبيل المثال، يمكن لأنماط التوزيع المتفاوتة أن تفسر لِمَ يقود البعض سياراتهم أفضل من المستوى المتوسط. فالكثيرون من الناس لديهم معدل حوادث يبلغ صفرا أو حادثا واحدا فقط أثناء حياتهم بأكملها. وبعض السائقين المتهورين لديهم الكثير، مما يرفع متوسط الحوادث إلى رقم أعلى بكثير مما يمر به معظم الناس. وينطبق هذا كذلك على توزيع الدخل: فمعظم الناس يكسبون أقل من المتوسط.

ما يمكنك فعله

خذ في حسبانك دائمًا معدلات التوزيع والمعدلات الأساسية. فالتحقق من الوسيط والمتوسط الحسابي، بالإضافة إلى المنوال (وهو القيمة الأكثر تكرارًا في التوزيع) سيساعدك على فهم البيانات. كما أن معرفة القيمة الأسمية سيسهل من وضع الأمور في سياقها، كما في مثال مرض التصلب المتعدد. وأخيرًا، استخدام المعدلات الطبيعية (واحد في كل مائة) طريقة أسهل بكثير لفهم القراء بدلاً من استخدام النسبة المئوية (1٪).

3- ما مدى موثوقية المعلومات؟

مشكلة حجم عينة البحث

جاء في تقرير مبني على دراسة مسحية نشرته صحيفة 'ديارو دي نابارا'، التي مقرها مدينة سرقسطة الإسبانية، أن "80٪ من الشعب غير راضين عن النظام القضائي." كيف يمكن الوصول لنتيجة كهذه من خلال عينة مكونة من 800 شخص من بين 46 مليون إسباني؟ هذا هراء بالتأكيد.

عند إجراء بحث على عدد ضخم من الجمهور (يربو على بضعة آلاف)، نادرًا ما تحتاج إلى أكثر من ألف شخص في العينة لتحقق هامش خطأ أقل من 3٪. وهذا يعني أنك إن أجريت الدراسة مرة أخرى مع عينة مختلفة تمامًا، ففي تسع حالات من بين كل عشرة ستكون الإجابات التي ستحصل عليها متقاربة بنسبة لا تتعدى 3٪. تتميز الإحصائيات بمكانة قوية كأداة للقياس، ومن النادر إلقاء اللائمة على حجم العينة التي تم مسحها عند إجراء استطلاعات رأي مزورة.

تناول الشاي يقلل من خطر الإصابة بالجلطة

لقد اعتدنا على المقالات حول فوائد تناول الشاي. ولا نستثني هذا الخبر القصير في صحيفة 'دي فيلت' الألمانية (http://bit.ly/welt–tea) الذي يقول إن الشاي يقلل من خطر الإصابة بالجلطة القلبية. وعلى الرغم من أن البعض يدرس بجدية آثار تناول الشاي، إلا أن الكثير من الأبحاث تفشل في الأخذ بعين الاعتبار

عوامل أنماط الحياة مثل النظام الغذائي أو المهنة أو ممارسة الرياضة.

في أغلب البلاد، يعتبر الشاي مشروب الطبقة العليا التي تهتم بصحتها. وإن لم يستقصِ الباحثون عوامل أنماط الحياة ويأخذونها في اعتبارهم في دراساتهم حول الشاي، فلن يخبرونا بشيء أكثر من أن "الأغنياء أكثر صحة وأنهم على الأغلب يشربون الشاي."

ما يمكنك فعله

مما لا شك فيه أن الرياضيات وراء معامل الارتباط وهوامش الخطأ في دراسات الشاي صحيحة بالتأكيد، على الأقل في أغلب الأحيان. ولكن إن لم يقم الباحثون بالبحث عن معامل الارتباط المعني (مثل: أن تناول الشاي يرتبط بممارسة الرياضة)، ستكون نتائجهم محدودة القيمة.

ولكونك صحافياً، فلن يكون من المنطقي أن تشكك في النتائج الرقمية لدراسة ما، مثل حجم العينة، إلا إذا كانت لديك شكوك قوية. ولكن سيسهل عليك رؤية إخفاق الباحثين في أخذ المعلومات ذات الصلة بعين الاعتبار.

— نيكولاس كايسر – بريل، شبكة Journalism++

نصائح للتعامل مع الأرقام في الأخبار

- أفضل نصيحة للتعامل مع البيانات هي أن تستمتع بعملك. فقد تبدو البيانات منيعة، ولكن إذا سمحت لها أن تخيفك فلن تحرز أي تقدم. تعامل معها على أنها لمجرد اللعب والاستكشاف، وستجد أنها غالبًا ما ستكشف أسرارها ودلالاتها بسهولة مفاجئة. لذا تعامل معها بنفس البساطة التي تتعامل بها مع أي أدلة أخرى، دون خوف أو تحيز. وعلى وجه الخصوص، اعتبرها تدريبًا على الخيال. كُن مبدعًا في التفكير في الأخبار البديلة التي يمكن أن تتسق مع البيانات واشرحها بشكل أفضل، ثم اختبرها مع المزيد من الأدلة. استعن بالسؤال القائل: 'أي تطور آخر يمكنه شرح ذلك؟' هو بمثابة دافع مفيد للتفكير في أن هذا الرقم – سواء أكان كبيرًا أم سيئًا، وسواء أستُخدِمَ كدليل واضح لهذا أو ذاك – قد لا يكون ما ظننته ذا مغزى للوهلة الأولى.

- لا تخلط بين الشك في البيانات والاستهانة بها. فالشك أمر جيد، أما الاستهانة فهي كالاستسلام والتخلي عن العمل. إذا كنتَ مؤمنًا بصحافة البيانات، وفي الغالب أنت كذلك، وإلا فما كنت لتقرأ هذا الكتاب، يجب أن تؤمن بأن البيانات تقدم شيئًا أفضل بكثير من الأكاذيب اللعينة لرسوم الكاريكاتير أو الحقائق القاتلة في عناوين الأخبار الصادمة التي تخطف الأبصار. إذا استُخدِمت البيانات بعناية، فإنها عادة ما تقدم لنا معرفة متبحرة. فلا يجوز أن نكون ساخرين أو ساذجين، بل يجب أن نكون منتبهين.
- إذا قلت لك إن معدل تناول المشروبات الكحولية قد ارتفع أيام الركود الاقتصادي، قد ترد بأن السبب في ذلك هو المعاناة من الاكتئاب. أما إذا قلت لك إن معدل تناول الكحول قد انخفض في الفترة ذاتها، فسترد بأن الجميع كانوا مفلسين. بعبارة أخرى، ما تقوله البيانات ليس له أي علاقة بالتفسير الذي تُصر على تقديمه، وهو أن الوضع سيء في جميع الأحوال سواء في حالة ارتفاع المعدل أو انخفاضه. والمقصود هو إن كنت تؤمن بالبيانات، فحاول أن تدعها تتحدث قبل أن تفرض مزاجك أو معتقداتك أو توقعاتك عليها. هناك الكثير من البيانات التي يمكنها تأكيد معتقداتك المسبقة إذا بحثت قليلاً. بعبارة أخرى، أرى أن صحافة البيانات، بالنسبة لي على الأقل، لا تضيف سوى قيمة قليلة إذا لم يكن الصحافي منفتح العقل، وهي موضوعية فقط بقدر سعيك لجعلها كذلك، وليس لمجرد اعتمادها على الأرقام.
- عدم اليقين أمر مقبول. فنحن نربط الأرقام بالمصدر والقناعة. وفي أغلب الأحيان، تكون الإجابة هي أنه لا توجد إجابة، أو أن الإجابة قد تكون أن هذه المعلومات هي أفضل ما لدينا ولكنها لا تزال بعيدة عن الدقة. أعتقد أن علينا الإفصاح عن هذه الأمور، وإذا بدا ذلك أشبه بطريقة فعالة للقضاء على الأخبار، فإن ردي سيكون أنها طريقة ممتازة لطرح أسئلة جديدة. وعلى نفس المنوال، يمكن عادة أن تكون هناك أكثر من طريقة منطقية لتحليل البيانات، فالأرقام لا تنحصر في كونها صحيحة أو خاطئة.
- يعتبر التحقيق خبرا في حد ذاته. فمن الممكن أن يكون سرد محاولاتك لاستكشاف الحقيقة عملاً صحفيًا رائعًا، حيث تتنقل من دليل إلى آخر لأن رقمًا

واحدًا لا يكفي للوقوف على الحقيقة، وهذا ينطبق على البيانات ذات الأحجام الكبيرة. وتتميز المصادر المختلفة بتقديم زوايا وأفكار جديدة وفهم أعمق. ولا أدري إذا كنا ما زلنا نتعلق كثيرًا بأن نكون في موضع صاحب المعرفة الذي يجب أن يحظى بالثقة حتى نجيب على أسئلة الجمهور، وعليه نفوت على أنفسنا الفرصة لإظهار كيف تم تحري الموضوع وسير عملية التحقيق.

- أفضل الأسئلة أقدمها: هل هذا فعلاً رقمٌ كبير؟ ومن أين أتى؟ هل أنت متأكد من أنه يصف ما تظن؟ هذه عموما مجرد دوافع للتفكير في البيانات والأمور المنزوية في الأركان التي استخرجناها بالنظر إلى رقم واحد، أو تعقيدات الواقع، أو المجموعة الواسعة من المقارنات المحتملة الأخرى على فترة زمنية أو فئة من الناس أو منطقة جغرافية، أو باختصار؛ السياق.

— مايكل بلاستلاند، صحفي مستقل

الخطوات الأساسية للعمل مع البيانات

عندما تبدأ في مشروع بيانات، ستكون في حاجة إلى استيعاب ثلاثة مفاهيم رئيسية على الأقل وهي:

- ينبغي لطلبات البيانات أن تبدأ بقائمة من الأسئلة تريد الإجابة عليها.
- عادة ما تكون البيانات فوضوية وتحتاج إلى التنظيم.
- قد تكون للبيانات خواص غير موثقة لايمكن التأكد من دقتها.



نموذج 5–2. بيانات فوضوية

كُن على علم بالأسئلة التي تود الإجابة عليها

العمل مع البيانات شبيه بإجراء مقابلة على الهواء. فأنت تطرح الأسئلة على البيانات وتجعلها تفصح عن الإجابات. وكما الحال في الضيف الذي يمكنه فقط الإجابة عن الأسئلة التي لديه معلومات عنها، لا يمكن لمجموعة البيانات أن تجيب سوى عن الأسئلة التي تمتلك سجلات صحيحة ومتغيرات مناسبة خاصة بها. وهذا يعني أنه ينبغي لك التفكير بدقة في الأسئلة التي تود الحصول على إجابات لها حتى قبل أن تحصل على بياناتك. وهذا يعني أنك في الأساس تعمل بشكل عكسي. أولاً، ضع قائمة بالعبارات المدعومة بأدلة من البيانات التي تود أن تنشرها في الخبر. ثم حدد أي المتغيرات والأرقام التي يجب أن تحصل عليها وتحللها حتى تتمكن من نشر هذه العبارات.

فكِّر في مثال يتعلق بتقارير الجرائم المحلية. فلنقل إنك تريد أن تكتب خبرًا عن أنماط الجرائم في مدينتك، وما تريد نشره من تعليقات يتعلق بعدد مرات وقوع الجريمة يوميًا وبالأيام التي غالبًا ما يحدث فيها مختلف أنواع الجرائم، بالإضافة إلى أي مناطق المدينة تعتبر بؤرة لفئات الجرائم المختلفة.

ستدرك أن طلبك للبيانات يجب أن يتضمن تاريخ ووقت تسجيل كل جريمة ونوعها (قتل أو سرقة أو سطو، إلخ) بالإضافة إلى عناوين مواقع حدوث الجريمة. إذن فالتاريخ والوقت والفئة والعنوان هي الحد الأدنى للمتغيرات التي تحتاجها للإجابة عن تلك الأسئلة.

ولكن لا بد من الانتباه إلى أن هناك عدداً من الأسئلة التي يحتمل أن تكون مثيرة للاهتمام، والتي لا تستطيع مجموعة البيانات ذات المتغيرات الأربعة أن تجيب عليها، مثل أعراق الضحايا وجنسهم، أو مجموع قيمة الممتلكات المسروقة، أو أي من ضباط الشرطة أكثر نشاطاً في عمليات الاعتقال. كما أنك قد تستطيع الحصول على سجلات لفترة زمنية معينة فقط، على سبيل المثال السنوات الثلاث الماضية، مما يعني أنك لن تتمكن من معرفة ما يتعلق بمتغيرات أنماط الجرائم أو ثباتها خلال فترة أطول. قد تكون تلك الأسئلة خارجة عن الإطار الذي تخطط له في الخبر، ولا بأس بذلك، ولكن لا ينبغي أن تبدأ بتحليل للبيانات ثم تقرر فجأة أنك مثلا في حاجة إلى معرفة نسبة الجرائم التي وقعت في مناطق مختلفة من

المدينة والتي تم حلها بالقبض على مرتكبيها.

أحد الدروس المستفادة هنا أنه عادة ما يكون من الأفضل طلب ’جميع‘ المتغيرات والسجلات في قاعدة البيانات، بدلاً من طلب مجموعة جزئية تجيب عن أسئلة الخبر المباشر. (في الواقع، قد يكون الحصول على جميع البيانات أقل كلفة من الحصول على مجموعة جزئية، هذا إن كان عليك دفع الرسوم للوكالة من أجل البرمجة الضرورية لاستخراج مجموعة جزئية). فيمكنك دائمًا أن تستخرج المجموعة الجزئية بنفسك، كما أن الحصول على مجموعة البيانات الكاملة سوف يمكنك من إجابة أسئلة جديدة قد تظهر أثناء إعدادك للتقرير أو حتى طرح أفكار جديدة من أجل تقارير المتابعة. قد تعني قوانين السرية أو السياسات الأخرى أن بعض المتغيرات، مثل هويات الضحايا أو أسماء المخبرين السريين، لا يمكن الإعلان عنها. ولكن قاعدة البيانات الجزئية أفضل من عدمها، ما دمت قد استطعت معرفة أي الأسئلة يمكن أو لا يمكن لقاعدة البيانات المنقحة أن تجيب عليها.

تنظيم البيانات الفوضوية

من أكبر مشاكل عمل قواعد البيانات هي أنك عادة ما تستخدم من أجل التحليل بيانات تم جمعها لأسباب بيروقراطية. المشكلة هي أن معيار الدقة لهذين السببين مختلف تماماً.

من أهم أدوار قاعدة بيانات نظام العدالة الجنائية مثلاً ضمان إحضار المتهم جونز من السجن ليمثل أمام القاضي سميث أثناء جلسة الاستماع. ولهذا السبب، لا يهم كثيراً إذا لم يكن تاريخ ميلاد جونز صحيحاً، أو إذا كان هناك خطأً إملائي في عنوانه، أو حتى إذا كان الحرف الأول من اسمه الأوسط خطأً. فعموماً، لا يزال بإمكان النظام أن يستخدم هذا السجل الناقص ليُحضر جونز إلى محكمة سميث في الميعاد المحدد.

ولكن مثل تلك الأخطاء يمكنها أن تشوه محاولات صحفي البيانات لاكتشاف الأنماط في قاعدة البيانات. ولهذا السبب، فإن أول وأكبر عمل عليك أن تتولاه عند الحصول على قاعدة بيانات جديدة هو أن تفحص مدى فوضويتها، ومن ثم القيام بتنظيمها. هناك طريقة جيدة وسريعة لفحص الفوضوية وهي أن تضع جداول

التوزيع التكراري لمتغيرات الفئات التي يتوقع أن يكون بها عدد صغير نسبياً من القيم المختلفة. (عندما تستخدم برنامج إكسل مثلاً، يمكنك فعل ذلك باستخدام جداول Filter أو Pivot لكل من المتغيرات الفئوية).

خُذ خانة ’الجنس‘ على سبيل المثال. قد تكتشف أن خانة الجنس تتضمن أياً من القيم المختلطة مثل: ذكر، أنثى، ذ، أ، 1، 0، ’ذكر‘، ’أنثى‘، إلخ، بما في ذلك الأخطاء الإملائية مثل ’أنث‘. وللقيام بتحليل سليم لخانة الجنس، عليك أن تضع معياراً – أي أن تختار ’ذ‘ و’أ‘ مثلاً – ثم تغير كل المتغيرات لتتماشى مع المعيار الجديد. ومن قواعد البيانات المعروفة الأخرى والتي لديها مثل تلك المشاكل سجلات تمويل الحملات الانتخابية الأمريكية، حيث يمكن أن يُكتب في خانة الوظيفة ’محام‘ أو ’وكيل‘ أو اختصار لكلمة ’وكيل‘ أو ’مستشار قانوني‘ أو ’محامي مرافعات‘ وأي من المتغيرات أو الأخطاء الإملائية؛ ومجددا فإن الحيلة هي وضع معيار لأسماء الوظائف في قائمة قصيرة تتضمن الاحتمالات.

وقد يثير تنظيم البيانات مزيدا من المشاكل عند العمل مع الأسماء. هل ’جوزيف ت. سميث‘ و’جوزيف سميث‘ و’ج. ت. سميث‘ و’جوز. سميث‘ و’جو سميث‘ كلها تشير إلى الشخص ذاته؟ قد يتطلب الأمر النظر إلى متغيرات أخرى مثل العنوان أو تاريخ الميلاد، أو حتى بحثاً أعمق في سجلات أخرى لتتمكن من التحديد. ولكن تستطيع أدوات مثل Google Refine أن تجعل عملية التنظيم ووضع المعيار أسرع وأقل إملالاً.

البيانات غير النظيفة

بفضل قوانين السجلات العامة القوية في الولايات المتحدة، لا يمثل الحصول على البيانات هنا مشكلة كبيرة خلافاً لما هو الحال في عدة بلدان أخرى. ولكن بمجرد أن نحصل عليها، نواجه مشاكل التعامل مع البيانات التي تم جمعها لغرض بيروقراطي، وليس لغرض التحليل. عادة ما تكون البيانات ’غير نظيفة‘ وتحتوي على قيم دون معيار. تلقيت عدة مرات بيانات لا تتماشى مع تصميم الملف ولا مع قاموس البيانات الذي يرافقها. تصر بعض الوكالات على إعطائك البيانات في صيغ غريبة مثل ملفات PDF، والتي يجب أن يتم تحويلها،

مما يجعلك تقدّر الحصول على مجموعة بيانات بدون مشاكل.

—— ستيف دويغ، كلية والتر كرونكايت للصحافة، جامعة أريزونا

قد تكون للبيانات مواصفات غير موثقة

إن ما يدعى بقاموس البيانات هو بمثابة الأساس بالنسبة لأي قاعدة بيانات. فعادة، يخبرك هذا الملف (قد يكون نصاً أو PDF أو حتى جدولا بيانيا) بصيغة ملف البيانات (نص محدد أو نص ذو عرض ثابت أو إكسل أو dBase أو ما إلى ذلك)، وترتيب المتغيرات وأسماء ونوع البيانات لكل منها (سلسلة نصية أو عدد صحيح أو رقم عشري). ستستخدم تلك المعلومات لمساعدتك على استيراد ملف البيانات إلى برنامج التحليل الذي تنوي استخدامه (إكسل، أكسيس، SPSS، جداول فيوجن، أي من برامج SQL أو ما إلى ذلك).

العنصر الأساسي الآخر في قاموس البيانات هو شرح أي رموز تستخدمها متغيرات معينة. فعلى سبيل المثال، يمكن أن يكون رمز الجنس هو '1=ذكر' و'0=أنثى'. ويمكن أن يرمز للجرائم حسب أرقام قانون الولاية القضائية لكل نوع من الجرائم. وقد تستخدم سجلات العلاج بالمستشفيات أياً من مئات الأرقام المكونة من خمسة أعداد لتشخيص الحالة التي يعالج منها المريض. وبدون قاموس البيانات، يصعب أو قد يستحيل تحليل مجموعات البيانات بشكل سليم.

ولكن حتى مع استخدام قاموس البيانات، قد تكون هناك مشاكل. ومثال على ذلك ما حدث لصحافيي صحيفة ميامي هيرالد في فلوريدا منذ بضع سنوات عندما كانوا يحللون المعدلات المتغيرة للعقوبة التي كان يصدرها قضاة مختلفون للأشخاص الذين اعتقلوا بسبب القيادة تحت تأثير الكحول. حصل الصحافيون على سجلات الإدانة من هيئة المحكمة وحللوا الأرقام في ثلاثة متغيرات مختلفة للعقوبات في قاموس البيانات: مدة الحكم بالسجن ومدة الحبس رهن التحقيق ومبلغ الغرامة. تفاوتت تلك الأرقام بشكل كبير فيما بين القضاة، مما أعطى الصحافيين دليلاً لنشر خبر حول قسوة بعض القضاة وتساهُل البعض الآخر.

ولكن بالنسبة لكل قاض، لم يصدر أيٍّ منهم أحكاماً بالسجن أو الاعتقال أو الغرامة في حوالي 1–2٪ من الحالات. لذا تضمن الرسم البياني لأنماط الأحكام

من كل قاضٍ جزءاً صغيراً من الحالات ’بدون عقوبة‘، وكأنها فكرة مُستدركة. وحين تم طبع الخبر والرسم البياني، شكا القضاة بأن صحيفة هيرالد تتهمهم بخرق قانون الولاية الذي ينص على إنزال العقوبة على أي شخص يُدان بتهمة القيادة تحت تأثير الكحول.

لذا عاد الصحافيون إلى مكتب كاتب المحكمة الذي كان قد سلمهم ملف البيانات وسألوه عما تسبب في هذا الخطأ. فقيل لهم أن القضايا المعنية تضمنت متهمين معوزين تم اعتقالهم للمرة الأولى وتغريمهم، إلا أنهم لا يملكون المال. لذا حكم عليهم القضاة بأداء خدمة المجتمع مثل جمع القمامة من الشوارع. ثم اتضح أن القانون الذي يستلزم إصدار العقوبة قد تم إقراره وسريانه بعد هيكلة قاعدة البيانات. لذا علم جميع كتاب المحاكم أن الأصفار في متغيرات السجن والاعتقال والغرامة بقاعدة البيانات تشير إلى خدمة المجتمع. ولكن هذا ’لم يكن‘ موجوداً في قاموس البيانات، وبالتالي اضطرت الصحيفة أن تنشر تصحيحاً لاحقاً.

الدرس المستفاد هنا هو أن تسأل دائماً الوكالة التي توفر لك البيانات عما إذا كانت هناك عناصر غير موثقة، أو إذا كانت رموز جديدة لم يتم ضمها إلى قاموس البيانات، أو عن التغييرات في شكل الملفات أو صياغتها أو عن أي أمر آخر. كما يجب أن تفحص دائماً نتائج تحليلك وتتساءل: "هل هذا منطقي؟" كان صحافيو صحيفة هيرالد يعدون رسماً بيانياً بموعد نهائي وركزوا كثيراً على متوسط مستويات العقوبة لكل قاض حتى أنهم فشلوا في الانتباه إلى القضايا القليلة التي لم تنته بعقوبات. كان عليهم أن يسألوا أنفسهم ما إذا كان من المنطقي أن يظهر كل القضاة كأنهم يخرقون قانون الولاية، حتى وإن كان هذا بدرجة ضئيلة للغاية.

— ستيف دويغ، كلية والتر كرونكايت للصحافة، جامعة أريزونا

بيانات مختلطة ومخبأة وغائبة

أذكر موقفاً طريفاً حدث عندما حاولنا الحصول على البيانات المجرية من دعم المزارع في الاتحاد الأوروبي: كانت كلها موجودة، ولكن مشتتة في صيغ تجاوزت كثيراً ملفات PDF ومختلطة ببيانات دعم المزارع الوطنية. اضطر مبرمجونا إلى العمل لأربع ساعات قبل أن نستطيع الاستفادة من تلك البيانات.

وقد قضينا أيضاً وقتا مثيرا للاهتمام مع بيانات دعم مصايد الأسماك في الاتحاد الأوروبي، التي تلتزم بنشرها وكالات الدفع الوطنية في الدول الأعضاء السبع وعشرين كلها. وفيما يلي مقتطف من تقرير أعددناه عن الموضوع (http: / / bit.ly /alfter–eu27): "في المملكة المتحدة مثلاً، تختلف صيغة البيانات ما بين صفحات البحث بصيغة HTML التي يسهل استخدامها، وملفات بصيغة PDF أو حتى قوائم بصيغ مختلفة مختبئة في أسفل التصريحات الصحفية. وكل هذا داخل دولة واحدة من الدول الأعضاء. وفي غضون ذلك، يتم نشر قوائم خالية في ألمانيا وبلغاريا. توجد بها العناوين المناسبة ولكن بدون أية بيانات."

— بريجيت ألفتر، Journalismfund.eu

رغيف الخبز البالغ سعره 32 جنيها استرلينيا

ذكر خبر في صحيفة 'ويلز يوم الأحد' كم تنفق حكومة ويلز على الوصفات الطبية للمنتجات الخالية من الغلوتين، وقد جاء في العنوان (/http://bit.ly walesonline–gluten–free) أن الحكومة تدفع 32 جنيها استرلينيا مقابل كل رغيف خبز. ولكنها كانت في الحقيقة 11 رغيفا يبلغ سعر كل واحد منها 2.82 جنيها.

ذكرت الأرقام التي جاءت في رد مكتوب من الجمعية الوطنية الويلزية وإحصائيات خدمة التأمين الصحي الوطنية NHS تكلفة كل وصفة طبية. ولكنهم لم يوفروا أي تعريف إضافي في قاموس البيانات لما قد تعنيه الوصفة الطبية أو كيف يمكن لخانة منفصلة تحدد الكمية أن تعرّفها.

كان الافتراض أن الوصفة تشير إلى شيء فردي كرغيف الخبز مثلاً، بدلاً مما كانت تشير إليه فعلاً وهي مجموعة مكونة من عدة أرغفة.

لم يهتم أي ممن أرسلوا الرد المكتوب أو من المكتب الإعلامي بمسألة الكمية حتى يوم الإثنين الذي تلا نشر الخبر.

لذا لا تفترض أن الملاحظات المتعلقة بخلفية المعلومات الموضحة لبيانات الحكومة ستساعد في إيضاح أي معلومات يتم تقديمها، أو أن المسؤولين عن البيانات سيدركون أن البيانات غير واضحة حتى وإن أبلغتهم بافتراضك الخاطئ.

تريد الصحف عموماً ما يشكل عناوين جيدة، لذا إن لم يكن هناك ما يتناقض

بشكل واضح مع التفسير المنشور، فعادة ما يكون من الأسهل اختيار ما يشكل عنواناً مثيراً وألا تتحقق بشكل دقيق حتى لا تخاطر بانهيار الخبر برمته، وخاصة في وجود موعد نهائي للنشر.



WalesOnline.co.uk

News from Wales and beyond, plus politics, business, sport, life and style... it's all here

Home | News | Rugby | Sports | Football | Lifestyle | Business | Classifieds

Health Check Wales | CardiffOnline | Vouchers

Hot Topics » Rebecca Aylward | Mike Phillips | Fracking | Gavin Henson

Home › News › Wales News

Prescriptions for gluten-free bread costing Welsh taxpayers £32

by Claire Miller, Wales On Sunday | Jul 17 2011

Like 24 | Tweet 5 | Share 7 | Email | Print

The Welsh NHS is forking out £32 a time for prescriptions for gluten-free bread.

The average prescription for the specialist food cost £32.27, and was provided to people with the serious condition coeliac disease.

Gluten-free loaves are available in shops for just £2.25, but are just one of scores of prescription items for which the Welsh NHS pays vast sums.

Related Tags

bread, gluten free, hayfever, nhs, over the counter medicines, painkillers, pasta, prescriptions, wales

(What's this)

نموذج 5–3. الوصفات الطبية للخبز الخالي من الغلوتين تكلف دافعي الضرائب في ويلز 32 جنيها (WalesOnline)

ولكن على الصحافيين أن يتحملوا مسؤولية التحقق من الادعاءات السخيفة حتى وإن كان ذلك يعني سقوط الخبر إلى أسفل قائمة الأخبار.

—— كلير ميلر، WalesOnline

ابدأ بالبيانات واختتم بالخبر

حتى تجذب انتباه القراء عليك أن تصدمهم برقم في عنوان الخبر يجعلهم ينتبهون ويعيرونك اهتمامهم. كما عليك أن تتمكن تقريباً من قراءة الخبر دون إدراك

أنه صادر من مجموعة بيانات. اجعله مثيراً وتذكر من هم جمهورك.

يمكنك أن تجد مثالاً على ذلك في مشروع قام به ’مكتب الصحافة الاستقصائية‘ باستخدام ’نظام الشفافية المالية‘ (http://bit.ly/ec–fts) لدى مفوضية الاتحاد الأوروبي، لتحديد ما أنفقته المفوضية، حيث تم بناء الخبر بالتعامل مع مجموعة البيانات مع وجود أسئلة معينة في الحسبان.

وقد بحثنا في البيانات عن كلمات دالة مثل ’كوكتيل‘ و’جولف‘ و’أيام في الخارج‘، مما سمح لنا بتحديد ما أنفقته المفوضية على تلك الأمور وبطرح عدة أسئلة وأخبار لنتابعها لاحقاً.

ولكن المصطلحات الدالة لا تعطيك دائماً ما تريده، أحياناً عليك أن تفكر فيما تطلبه حقاً. أثناء هذا المشروع أردنا أيضاً أن نعرف كم ينفق المفوضون على السفر بطائرات خاصة ولكن لأن مجموعة البيانات لم تتضمن كلمات مثل ’طائرة خاصة‘ اضطررنا للحصول على أسماء وكلاء سفرهم بطرق أخرى. وبمجرد أن وجدنا اسم وكيل السفر لدى المفوضين، شركة ’Abelag‘، تمكنّا من الاستفسار من البيانات لمعرفة كم يُنفَق على خدماتها.

وبهذا النهج كان لدينا هدف واضح ومحدد عند سؤال البيانات، وهو أن نعثر على رقم يزودنا بعنوان للخبر يتبعه اختيار اللون.

هناك نهج آخر للتناول، وهو البدء بقائمة سوداء والبحث عن الاستثناءات، ومن الطرق السهلة للحصول على خبر من البيانات أن تدرك ما لا يجب أن تجده هناك! وثمة مثال جيد على كيفية النجاح في ذلك قدمه مشروع تعاوني للصناديق الهيكلية الأوروبية بين صحيفة الفاينانشيال تايمز ومكتب الصحافة الاستقصائية.

سألنا البيانات حسب قوانين المفوضية حول نوع الشركات والمؤسسات التي يجب منعها من تلقي الدعم من الصناديق الهيكلية. ومثال على ذلك الإنفاق على التبغ ومنتجيه.

وعبر سؤال البيانات عن أسماء شركات التبغ ومنتجيه ومزارعيه، وجدنا بيانات أظهرت أن ’شركة التبغ البريطانية الأمريكية‘ تلقت مليونا ونصف مليون يورو تسلمها مصنع لها في ألمانيا.

وبما أن التمويل كان يخالف قوانين المفوضية للإنفاق، فقد كانت هذه طريقة

سريعة للعثور على خبر في البيانات.

لا يمكنك أن تعرف أبداً ما ستجده في مجموعة البيانات، لذا فقط الق نظرة متفحصة. عليك أن تكون جريئاً جداً وعادة ما ينجح هذا الأسلوب بشكل أفضل عند محاولة تحديد خصائص واضحة تظهر من خلال الفرز (وفقاً للأكبر أو الأكثر تطرفاً أو الأكثر انتشاراً، إلخ).

—— كايلاين بار، مجموعة Citywire

الأخبار المبنية على البيانات

يمكن لصحافة البيانات أن تعطينا أحياناً انطباعاً بأنها تتعلق أساساً بعرض البيانات، كالتوضيحات المرئية التي تنقل بسرعة وبقوة فهماً لجانب من جوانب الأرقام أو قواعد البيانات التفاعلية التي تتيح للأفراد البحث عن شارعهم أو المستشفى المحلي مثلاً. يمكن لكل هذا أن يكون قيّما جداً، ولكن مثله مثل أشكال أخرى للصحافة، إذ يجب على صحافة البيانات أيضاً أن تتعلق بالأخبار. إذن فما هي الأخبار التي يمكن العثور عليها في البيانات؟ بحسب خبرتي في شبكة بي. بي. سي، وضعت قائمة أو 'تصنيفاً' للأنواع المختلفة من الأخبار المبنية على البيانات.

أعتقد أنه من المفيد أن تضع القائمة أدناه في الحسبان، ليس فقط عندما تحلل البيانات، ولكن أيضاً في المرحلة التي تسبق ذلك، عند جمعها (سواء أكنت تبحث عن مجموعات بيانات متاحة للعامة أم تجمع طلبات حرية المعلومات).

1- القياس

أبسط الأخبار؛ العد أو الحساب "أنفقت المجالس المحلية في جميع أنحاء البلاد ما مجموعه كذا مليار جنيه استرليني على مشابك الورق العام الماضي." ولكن غالباً ما يصعب معرفة ما إذا كان ذلك كثيراً أم قليلاً. ولذلك تحتاج إلى السياق، والذي يمكن أن يوفره الآتي:

2- التناسب

"أنفقت المجالس المحلية في العام الماضي ثلثي ميزانياتها للأدوات المكتبية على مشابك الورق."

3– المقارنة الداخلية

"أنفقت المجالس المحلية على مشابك الورق أكثر مما أنفقت على توصيل الوجبات لمنازل كبار السن."

4– المقارنة الخارجية

"بلغ إنفاق المجالس المحلية على مشابك الورق العام الماضي ضعف ميزانية الدولة للمساعدات الخارجية."

توجد طرق أخرى لاستكشاف البيانات بالسياق أو المقارنة:

5– التغيّر عبر الزمن

"ازداد إنفاق المجالس المحلية على مشابك الورق ثلاثة أضعاف خلال الأربعة أعوام الماضية."

6– "جداول المقارنة"

عادة ما تكون تلك الجداول جغرافية أو حسب المؤسسة، وعليك أن تتأكد من أن أساس المقارنة عادل (مثل أخذ عدد السكان المحليين في الحسبان).

"ينفق مجلس بورسيتشاير على مشابك الورق لكل موظف أكثر مما تنفق أي جهة محلية أخرى، وبمعدل أكثر من المتوسط القومي بأربعة أضعاف."

أو يمكنك تقسيم مواضيع البيانات إلى مجموعات:

7– التحليل حسب التصنيف

"تنفق المجالس التي يديرها الحزب البنفسجي على مشابك الورق 50٪ أكثر من تلك التي يديرها الحزب الأصفر."

أو يمكنك الربط بينهما بالأرقام:

8– بعلاقة ربط

"المجالس التي يديرها سياسيون تلقوا تبرعات من شركات الأدوات المكتبية تنفق أكثر على مشابك الورق، وبزيادة في الإنفاق بمتوسط 100 جنيه لكل جنيه تم التبرع به."

ولكن بالطبع عليك أن تتذكر أن الارتباط والسببية ليسا الأمر ذاته.

إذن فإن كنت تتحقق من الإنفاق على مشابك الورق، هل تحصل أيضاً على الأرقام التالية:

- إجمالي الإنفاق لتعطي سياقاً مناسباً؟
- تحليل جغرافي أو تاريخي أو غير ذلك من التفاصيل لتعطي بيانات للمقارنة؟
- البيانات الإضافية التي تحتاجها لضمان عدالة المقارنات، كعدد السكان مثلاً؟
- بيانات أخرى يمكنها أن تعطي تحليلاً مثيراً للاهتمام لمقارنتها أو الربط بينها وبين الإنفاق؟

—— مارتن روزنبام، بي. بي. سي.

صحافيو البيانات يناقشون اختياراتهم من الأدوات

ستطلق بياناتك صفيراً حاداً عند خروجها من غلافها المحكم. فما العمل الآن؟ ما الذي تبحث عنه؟ وما هي الأدوات التي تلزمك حتى تنغمس في العمل؟ لقد طلبنا من صحافيي البيانات الاستطراد قليلاً في الحديث عن كيفية عملهم مع البيانات. وفيما يلي ردودهم:

—— ليزا إيفانز، صحيفة الغارديان

"نحن في مدونة بيانات الغارديان نحب كثيراً أن نتفاعل مع قرائنا وأن نتيح لهم نسخ صحافة البيانات لدينا مما يعني أن بإمكانهم الإضافة إلى ما نقوم به وأحياناً يلاحظون ما لم نلاحظه نحن. لذا فكلما كانت أدوات البيانات أسهل في الاستخدام بشكل تلقائي، كان ذلك أفضل. نحن نحاول أن نختار أدوات يمكن لأي شخص أن يعتاد عليها بدون تعلم لغة برمجة أو الحصول على تدريب خاص وبدون مصاريف باهظة.

ولهذا السبب نستخدم حالياً منتجات جوجل بشكل مكثف. فكل مجموعات البيانات التي ننظمها وننشرها متاحة في جداول جوجل البيانية مما يعني أن من لديهم حسابات جوجل يمكنهم تحميل البيانات واستجلابها إلى حسابهم الخاص

لصنع أشكال بيانية خاصة بهم وتصنيفها ووضع جداول محورية، أو بإمكانهم استيراد البيانات إلى أي أداة من اختيارهم.

أما عند وضع خرائط للبيانات، فنستخدم جداول جوجل فيوجن. وعندما نضع خرائط ملونة للبيانات باستخدام فيوجن، نشارك ملفات KML الخاصة بنا لنتيح للقراء تحميلها ووضع خرائط ملونة خاصة بهم وربما حتى إضافة طبقات إضافية من البيانات إلى الخريطة الأصلية في مدونة البيانات. ومن الخصائص الإيجابية الأخرى لتلك الأدوات من جوجل أنها تعمل في العديد من المنصات التي يستخدمها قراؤنا للدخول إلى المدونة، مثل الكمبيوتر التقليدي والهاتف المحمول والكمبيوتر اللوحي.

وبالإضافة إلى جداول جوجل البيانية وجداول فيوجن، نستخدم أداتين أخريين في عملنا اليومي. أولهما Tableau، للتوضيح المرئي لمجموعات البيانات متعددة الأبعاد، والثانية هي ManyEyes من أجل التحليل السريع للبيانات. وبما أن أيٍ من تلك الأدوات ليست مثالية تماما، نواصل البحث عن أدوات توضيح مرئي أفضل ليتسنى لقرائنا الاستفادة منها."

—— سينثيا أومورتشو، صحيفة فاينانشال تايمز

"هل سأتمكن يوما ما من أن أكون مبرمجاً يقوم بعملية الترميز اي يكتب الكود؟ هذا مستبعد تماماً! أنا بالتأكيد لا أعتقد أن على كل الصحافيين أن يعرفوا كيفية كتابة الكود. ولكنني أعتقد أنه من المهم بالنسبة لهم أن يتمتعوا بإدراك عام لما هو ممكن وأن يعرفوا كيفية التحدث إلى المبرمجين.

إذا كنت مبتدئاً، فابدأ تدريجياً متمهلا ولا تتسرع. عليك أن تقنع زملاءك ومحرريك أن العمل مع البيانات يمكنه جلب الأخبار التي لم تكن لتحصل عليها بدون البيانات، وأن الأمر يستحق. وبمجرد أن يروا أهمية هذا النهج، يمكنك التوسع إلى أخبار ومشاريع أكثر تعقيداً.

أنصحك بتعلم استخدام برنامج إكسل والبدء ببعض الأخبار البسيطة أولاً. أبدأ على نطاق ضيق واعمل لتتوسع وصولاً إلى تحليل قواعد البيانات وتخطيطها. يمكنك عمل الكثير في إكسل، فهو يعتبر أداة قوية للغاية ومعظم الناس لا يستخدمون سوى جزءٍ صغير من وظائفه. إذا استطعت، انضم إلى دورة تدريبية

لبرنامج إكسل من أجل الصحافيين مثل التي يقدمها مركز الصحافة الاستقصائية.

أما بخصوص تفسير البيانات: لا تستخف بهذا الأمر. يجب أن تراعي الدقة والحرص. انتبه للتفاصيل وشكك في نتائجك. اكتب ملاحظات حول كيفية معالجتك للبيانات واحتفظ بنسخة من البيانات الأصلية. فمن السهل أن ترتكب خطأ، فأنا دائماً أقوم بالتحليل مرتين أو ثلاث مرات بالعودة إلى الصفر تقريباً. ومن الأفضل أن تطلب من المحرر أو أي شخص آخر أن يحلل البيانات بمعرفته وبشكل منفصل ويقارن النتائج."

—— سكوت كلاين، بروبابليكا

"تعتبر القدرة على كتابة ونشر برامج الكمبيوتر بنفس السرعة التي يستطيع بها الصحفي كتابة الخبر أمرا حديثا. فقد كان الأمر يستغرق وقتا أطول. ولكن الأمور تغيرت بفضل تطوير إطارين للتطور السريع ذي المصدر الحر/المفتوح: وهما Django وRuby on Rails الذين تم إطلاقهما للمرة الأولى في أواسط العقد الأول من الألفية.

لقد طوّر إطار Django – الذي كُتب بلغة البرمجة بايثون – آدريان هولوفاتي وفريق يعمل في غرفة أخبار صحيفة لورنس جورنال وورلد التي تصدر في مدينة لورنس بولاية كانساس. أما إطار Ruby on Rails فقد طُوّر في شيكاغو من قِبَل دايفيد هاينماير وشركة 37Signals لتطبيقات الإنترنت.

وعلى الرغم من أن كلا الإطارين يتخذ نهجاً مختلفاً في 'أنماط MVC'، إلا أن كليهما ممتاز ويتيح بناء تطبيقات بالغة التعقيد بسرعة فائقة. فهما يزيلان بعض العمل البدائي في بناء التطبيقات. أما أمور مثل وضع أو جلب أشياء من قاعدة البيانات ومطابقة عناوين URL مع كود معين في التطبيق فيتم وضعها في الإطُر الرئيسية حتى لا يضطر المطورون إلى الترميز أو كتابة الكود من أجل القيام بمثل تلك الأمور الأساسية.

وفي حين لا توجد دراسة رسمية عن فرق تطبيقات الأخبار في الولايات المتحدة، هناك فهم عام بأن أغلب الفرق تستخدم أحد هذين الإطارين لبناء تطبيقات الأخبار المدعومة من قاعدة البيانات. وهنا في بروبابليكا نستخدم إطار Ruby on Rails.

إن تطوير خدمات الإمداد لخادم "سيرفر" الويب السريع ’سلايس‘ مثل خدمات أمازون قد خفف مما كان يبطئ من عملية نشر أي تطبيق للويب.

فضلاً عن ذلك، نحن نستخدم أدوات عادية للعمل مع البيانات، مثل Google Refine وإكسل لتنقية البيانات؛ وSPSS وR للإحصاءات؛ وArcGIS وQGIS لنظام المعلومات الجغرافي GIS، وبرنامج Git لإدارة كود المصدر؛ و TextMate وVim وSublime Text للترميز وكتابة الكود؛ ومزيج بين MySQL وPostgreSQL وSQL Server لقواعد البيانات. ولقد بنينا إطار جافاسكريبت خاصا بنا يُدعى "Glass" وهو يساعدنا على بناء تطبيقات ثقيلة باستخدام جافاسكريبت بسرعة فائقة."

—— شيريل فيليبس، صحيفة سياتل تايمز

"أحياناً تكون أفضل الأدوات هي أبسطها – فمن السهل الاستخفاف بقوة الجدول البياني. ولكن استخدام الجدول البياني عندما كان كل شيء يعمل بنظام تشغيل DOS مكنني من فهم الصيغة المعقدة لاتفاق الشراكة بين مُلاك فريق ’تكساس رينجرز‘ عندما كان جورج بوش أحد المُلاك الرئيسيين. يمكن للجدول البياني أن يساعدني على تحديد القيم المتطرفة أو الأخطاء في الحسابات. ويمكنني أن أكتب برنامجاً نصياً للتنقية وأكثر من ذلك. وهذا أساسي في صندوق أدوات صحافي البيانات. إضافة إلى ذلك، فإن أدواتي المفضلة لديها المزيد من القوة – مثل SPSS للتحليل الإحصائي وتخطيط البرامج التي تمكنني من رؤية الأنماط جغرافياً."

—— غريغور آيش، مؤسسة المعرفة المفتوحة

"أنا معجب جداً ببايثون، فهي لغة برمجة مفتوحة المصدر وتسهل قراءتها وكتابتها (فليس عليك مثلاً أن تكتب فاصلة منقوطة بعد كل سطر). والأهم من ذلك أن بايثون لديها قاعدة مستخدمين هائلة وبالتالي لديها ملحقات (أو باقات) لأي شيء تريده.

وأظن أن صحافيي البيانات نادراً ما يحتاجون إطار Django، فهو إطار

تطبيق إنترنت بلغة بايثون، أي أنه أداة لبناء تطبيقات إنترنت أكبر ومبنية على قواعد البيانات. وهذا بالتأكيد أفضل ما يمكن الحصول عليه للرسوم البيانية التفاعلية الصغيرة.

أستخدم أيضاً تطبيق QGis، وهي عدة أدوات مفتوحة المصدر توفر مجموعة واسعة من وظائف نظام المعلومات الجغرافي GIS التي يحتاجها صحافيو البيانات ممن يتعاملون مع البيانات الجغرافية بين حين وآخر. إذا احتجت إلى تحويل بيانات جغرافية مكانية من صيغة إلى أخرى، فأنت تحتاج إلى تطبيق QGis الذي يمكنه التعامل مع جميع صيغ البيانات الجغرافية الموجودة تقريبا (مثل Shapefiles أو KML أو GeoJSON إلخ). وإذا أردت استبعاد بعض المناطق، يمكن لهذا التطبيق أن يقوم بذلك أيضاً. بالإضافة إلى أنه توجد جماعة كبيرة من مستخدمي تطبيق QGis لذا يمكنك أن تجد الكثير من الموارد مثل الدروس التعليمية (http://bit.ly/geottingen–tutorial) على مواقع الإنترنت.

أما لغة R للبرمجة فقد وُضِعت أساساً لتكون أداة للتوضيح المرئي العلمي. ومن الصعب أن تجد أي أسلوب للتوضيح المرئي أو تقنية لمعالجة البيانات لم تتم كتابتهما بلغة R، فهي عالم قائم بذاته وتُعتبر قِبلة تحليل البيانات المرئية. ومن عيوبها أن عليك تعلُّم لغة برمجة أخرى. ولكنك بمجرد أن تأخذ الخطوة الأولى في التعلّم، ستجد أنه لا توجد أي أداة أقوى من R. يمكن لصحافيي البيانات المدرّبين استخدامها لتحليل مجموعات البيانات الهائلة التي تتجاوز حدود إكسل (مثلاً إن كان لديك جدول مكون من ملايين الصفوف).

ومن الأمور الجيدة في R أن بإمكانك تسجيل محضر دقيق 'بروتوكول' لما تفعله بالبيانات والعملية الكاملة بدءاً من قراءة ملف CSV وصولاً إلى الرسومات البيانية. إذا تغيرت البيانات، يمكنك وضع الرسم البياني بضغطة زر واحدة. وإذا تساءل أحدهم عن نزاهة رسمك البياني، يمكنك أن توضح بالضبط المصدر الذي يسمح للجميع بإعادة وضع الرسم البياني ذاته وحدهم (أو حتى لتكتشف أخطاءك).

إن NumPy + MatPlotLib يعتبران طريقة لفعل الشيء نفسه في لغة بايثون. وهذا خيار متاح إذا كنت بالفعل مُدَرباً تدريباً جيداً على استخدام تلك اللغة. كما

أنهما مثالان على باقات بايثون ويمكن استخدامهما لتحليل البيانات وتوضيحها مرئياً وكلاهما يقتصر على التوضيحات المرئية الثابتة، ولا يمكن استخدامهما لوضع رسوم بيانية تفاعلية مع تعريفات الأدوات والأمور الأكثر تطوراً.

أنا لا أستخدم MapBox، ولكنني سمعت أنه أداة ممتازة إذا أردت أن تُعِد خرائط أكثر تعقيداً حسب OpenStreetMap. فهو يسمح لك مثلاً بتخصيص شكل الخرائط (الألوان والأسماء وما إلى ذلك). وهناك أيضاً مرافق لـ MapBox وهو Leaflet الذي يعتبر أساساً مكتبة جافاسكريبت ذات مستوى أعلى للتخطيط الذي يتيح لك التنقل بسهولة بين مزودي الخرائط (مثل MapBox وخرائط جوجل وبينغ، إلخ).

RaphaelJS هي مكتبة توضيح مرئي منخفضة المستوى نوعاً ما وتتيح لك العمل بالأساسيات (كالدوائر والخطوط والنصوص) وتحريكها وجعلها تفاعلية. لا يوجد ما يشبه مخطط أعمدة بيانية فيه، لذا عليك أن ترسم مجموعة من المستطيلات بنفسك.

ولكن الأمر الجيد بخصوص Raphael هو أن كل ما تفعله سيعمل أيضاً باستخدام إنترنت إكسبلورَر (IE). وهذا لا يحدث مع عدة مكتبات (جيدة) أخرى للتوضيح المرئي مثل d3. وللأسف لا يزال العديد من المستخدمين يستخدمون IE ولا توجد غرفة أخبار تستطيع أن تتحمّل تجاهُل 30٪ من مستخدميها.

إلى جانب RaphaelJS، هناك أيضاً خيار إنشاء نظام Flash احتياطي لـ IE. وهذا أساساً ما تفعله صحيفة نيويورك تايمز، مما يعني أن عليك تطوير كل تطبيق مرتين.

ما زلت غير مقتنع بـ "أفضل" عملية لنقل التوضيح المرئي لـ IE وبرامج التصفح الحديثة. وغالباً ما أجد أن تطبيقات RaphaelJS قد تعمل ببطء شديد على IE، أبطأ بعشر مرات من عملها باستخدام فلاش على برامج التصفح الحديثة. لذا فنظام فلاش الاحتياطي قد يكون خياراً أفضل إن أردت توفير توضيحات مرئية متحركة ذات جودة عالية لكل المستخدمين."

— ستيف دويغ، كلية والتر كرونكايت للصحافة، جامعة أريزونا

"الأداة التي ألجأ إليها هي إكسل، وهو البرنامج الذي يمكنه التعامل مع أغلب مشاكل الصحافة بمساعدة الكمبيوتر CAR ومن مميزاته أنه سهل للتعلُّم ومتاح لمعظم الصحافيين. عندما أريد دمج الجداول، عادة ما أستخدم أكسيس، ولكنني بعد ذلك أصدّر الجدول الذي دمجته إلى إكسل للعمل عليه أكثر. وأستخدم ArcMap من ESRI للتحليلات الجغرافية، فهو قوي وتستخدمه الوكالات التي تجمع البيانات المرمزة جغرافياً. وTextWrangler رائع لفحص بيانات النصوص ذات التصاميم والمحددات الغريبة، ويمكنه أيضاً القيام بعملية بحث واستبدال متطورة للتعبيرات العادية. وعند الحاجة إلى تقنيات إحصائية مثل الانحدار الخطي، أستخدم SPSS؛ فلديه قائمة سهلة الاستخدام. أما بالنسبة للأمور الصعبة، مثل العمل مع مجموعات البيانات التي لديها ملايين السجلات مما قد يوجب تصفية جدية وتحولات مبرمجة للمتغيرات، فأستخدم برنامج SAS.

—— بريان بوير، صحيفة شيكاغو تريبيون

"من ضمن الأدوات التي اخترناها بايثون وDjango للبرمجة واستخراج البيانات والتعامل معها؛ كما اخترنا PostGIS وQGIS وعدة أدوات MapBox لبناء خرائط إنترنت متميزة. أما بالنسبة لـ R وNumPy + MatPlotLib، فهم يحاربون حالياً من أجل التفوق لنختارهم كعدة أدوات لتحليل البيانات التمهيدي، على الرغم من أن أداة البيانات المفضلة لدينا مؤخراً مصنوعة منزلياً، وهي CSVKit. كل ما نقوم به تقريباً ننشره على الشبكة السحابية."

—— أنجيليكا بيرالتا راموس، صحيفة لا ناسيون الأرجنتينية

"نحن في صحيفة 'لا ناسيون' نستخدم ما يلي:

- إكسل لتنظيم البيانات وترتيبها وتحليلها؛
- جداول جوجل البيانية للنشر والتواصل مع الخدمات مثل جداول جوجل فيوجن ومنصة البيانات المفتوحة من Junar؛
- Junar لمشاركة بياناتنا ودمجها في مقالاتنا وتدويناتنا؛
- Tableau Public للتوضيحات المرئية التفاعلية للبيانات؛

- Qlikview وهي أداة سريعة للغاية للمعلومات التجارية نستخدمها لتحليل مجموعات البيانات الضخمة وتصفيتها.
- NitroPDF لتحويل ملفات PDF إلى ملفات نصية وملفات إكسل؛
- جداول جوجل فيوجن للتوضيحات المرئية للخرائط."

— بيدرو ماركون، مبرمج في منظمة الشفافية الدولية

"بصفتنا جماعة شعبية بدون أي انحياز تقني، نحن في 'مبرمجي الشفافية' Transparency Hackers نستخدم الكثير من الأدوات ولغات البرمجة. لكل عضو مجموعة تفضيلاته الخاصة به، وهذا التنوع الكبير يعتبر نقطة قوتنا وضعفنا في نفس الوقت. فالبعض منّا يبني 'Transparency Hacker Linux Distribution' والذي يمكننا تشغيله في أي مكان دون تثبيت والبدء في برمجة البيانات. تحتوي الأدوات المتعددة تلك على بعض الأدوات المثيرة للاهتمام ومكتبات للتعامل مع البيانات مثل Refine وRStudio وOpenOfficeCalc (وهي أداة عادة ما يتجاهلها المحترفون، ولكنها مفيدة جداً للأمور الصغيرة والسريعة). نستخدم أيضاً Scraperwiki كثيراً لوضع النماذج الأولية وحفظ نتائج البيانات على الإنترنت.

أما بالنسبة للتوضيحات المرئية والرسوم البيانية للبيانات، فهناك العديد من الأدوات التي تحوز إعجابنا. بايثون وNymPy قويتان للغاية. عدد قليل من الجماعة من يستخدم R ولكن في النهاية ما زلت أعتقد أن مكتبات جافاسكريبت لتخطيط الرسوم البيانية مثل d3 وFlot وRaphaelJS تُستخدم في أغلب المشاريع. وأخيراً، كنا نقوم بتجارب كثيرة في التخطيط ووجدنا أن Tilemill أداة مثيرة للاهتمام."

استخدام التوضيح المرئي للبيانات لاستكشاف ما بها من أفكار

"إن التوضيح المرئي هام للغاية بالنسبة لتحليل البيانات، فهو خط الهجوم الأمامي الذي يكشف عن الهيكل الدقيق للبيانات التي لا يمكن استيعابها بأي طريقة أخرى. نحن نكتشف تأثيرات لا يمكن تصورها ونتحدى تلك التي يمكن تصورها."

— ويليام س. كليفلاند (من كتاب: التوضيح المرئي للبيانات – الناشر: Hobart Press)

إن البيانات في حد ذاتها والمكونة من "بت" و"بايت" ومخزنة في ملف على

القرص الصلب للكمبيوتر، غير مرئية. ولكي نتمكن من رؤية البيانات وفهمها، نحتاج إلى توضيحها بصرياً. في هذا الفصل سأستخدم مفهوماً أوسع لمصطلح 'التوضيح المرئي'، بما في ذلك حتى التمثيل النصي للبيانات. وعلى سبيل المثال، يمكن لتحميل مجموعة البيانات على برنامج للجداول البيانية أن يُعتبر توضيحاً مرئياً للبيانات. لذا فالسؤال يجب ألا يكون حول ما إذا كان الصحافيون يحتاجون إلى توضيح البيانات مرئياً أم لا، بل حول أي نوع من التوضيح المرئي يمكن أن يكون الأكثر إفادة وفي أية حالة.

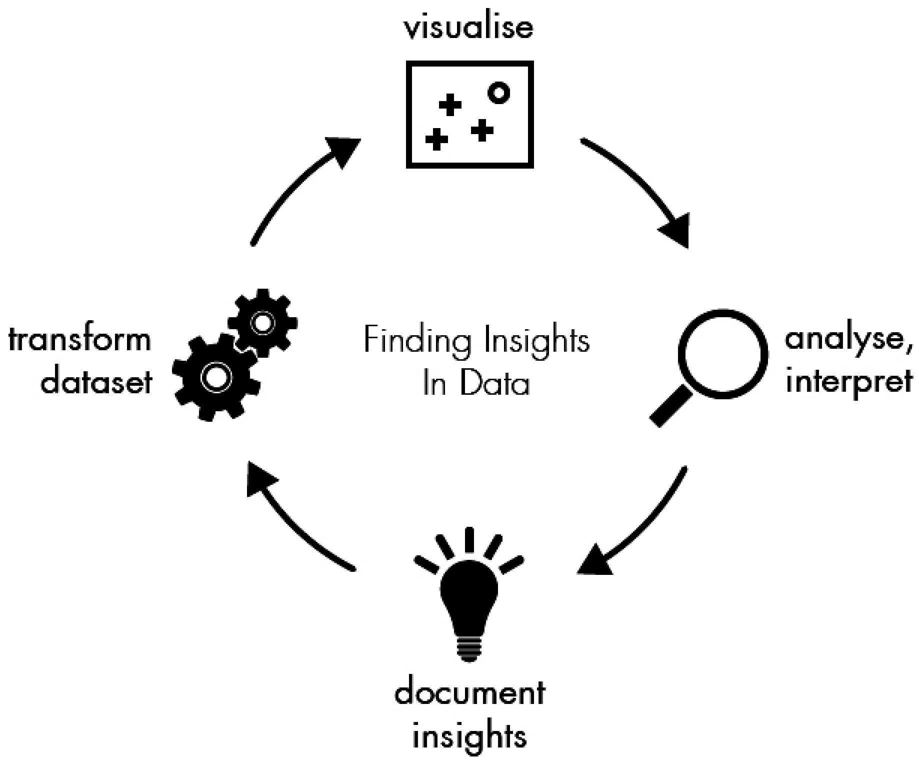
وبعبارة أخرى، متى يكون من المنطقي تجاوز التوضيح المرئي في شكل الجدول؟ الإجابة القصيرة هي: في جميع الأحوال تقريباً. فالجداول وحدها قطعاً لا تكفي لإعطائنا نظرة عامة على مجموعة بيانات، ولا تتيح لنا التعرف فوراً على الأنماط داخل البيانات. المثال الأكثر شيوعاً هنا هو الأنماط الجغرافية التي يمكن ملاحظتها فقط بعد توضيح البيانات مرئياً على الخريطة. ولكن هناك أيضاً أنواعاً أخرى من الأنماط التي سنراها لاحقاً في هذا الفصل.

استخدام التوضيح المرئي لاكتشاف الأفكار

من غير الواقعي أن نتوقع من أدوات وتقنيات التوضيح المرئي للبيانات أن تطلق وابلاً من الأخبار المكتملة من داخل مجموعات البيانات. فلا توجد قواعد ولا 'بروتوكول' يضمن لنا الخروج منها بأخبار. بدلاً من ذلك، أعتقد أنه من الأكثر منطقية أن نبحث عن 'الأفكار' التي يمكن نسجها بمهارة لتكوين أخبار على يد الصحافيين البارعين.

من المرجح أن يمنحنا كل توضيح مرئي جديد قدرا من البصيرة بخصوص بياناتنا. وقد تكون بعض هذه الأفكار معروفة بالفعل (ولكن ربما لم يتم إثباتها بعد) بينما قد تكون بعض الأفكار الأخرى جديدة تماماً أو قد يكون إدراكها مفاجأة لنا. قد تكون بعض الأفكار الجديدة بداية لخبر، بينما قد يكون البعض الآخر نتيجة لأخطاء البيانات والتي من المرجح العثور عليها عبر التوضيح المرئي للبيانات.

ولكي نجعل العثور على الأفكار في البيانات أكثر فعالية، أستعرض لكم العملية المذكورة أدناه في نموذج 5–4 (وفيما تبقى من هذا الباب) والتي أظن أنها مفيدة جداً:



نموذج 5–4. أفكار البيانات: التوضيح المرئي (غريغور آيش)

سنناقش كل واحدة من هذه الخطوات بالتفصيل في هذا القسم

تعلّم كيفية توضيح البيانات مرئياً

يزودنا التوضيح المرئي بمنظور فريد فيما يخص مجموعة البيانات. ويمكنك أن توضح البيانات بصرياً بعدة طرق.

تعتبر الجداول وسيلة إيضاحية فعالة للغاية عند التعامل مع عدد محدود نسبياً من النقاط محل الاهتمام الموضحة ببيانات وأرقام، فالجداول تظهر الأسماء والمبالغ/الكميات بشكل غاية في النظام والترتيب وتكشف عن كامل إمكانياتها الكامنة عند استخدامها ضمن قدرات فرز البيانات وتصفيتها. علاوة على ذلك، اقترح إدوارد توفتي إدراج رسوم بيانية مبسطة ضمن أعمدة الجدول، مثلاً شريط واحد في كل صف أو رسم بياني خطي صغير (والذي عُرف منذ ذلك الحين باسم

sparkline). ولكن كما ذكرنا في المقدمة، لا تزال إمكانات الجداول محدودة، فهي ممتازة في إظهار الخطوط العريضة أو الصورة الإجمالية أحادية البعد مثل العشرة الأوائل، ولكنها ضعيفة فيما يتعلق بمقارنة الأبعاد المتعددة في نفس الوقت (مثل تعداد السكان لكل دولة عبر فترة من الزمن).

Major Groups	Apr-07	—	Apr-08	% YoY	% MoM	%Wt
Food	2,532		2,588	2.2	-0.3	40.9
Hospitality & Service Industries	1,159		1,195	3.1	0.1	18.7
Household Goods	858		933	8.7	0.5	13.9
Other, Pharma, Watches	552		625	13.3	-0.8	8.9
Department Stores	482		500	3.8	0.1	7.6
Clothing & Soft Goods	421		453	7.6	0.3	6.8
Recreational Goods	190		222	16.5	0.4	3.1
Total Retail Sales	6,194.2		6,313.1	5.2	-0.1	

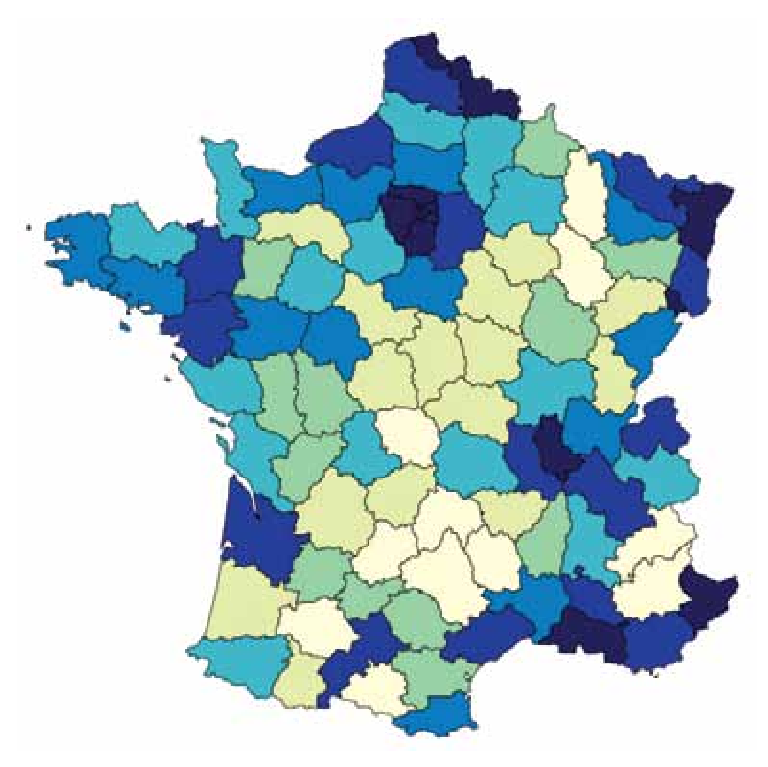
Group	Apr-07	—	Apr-08	% YoY	% MoM	%Wt
Food						
Supermarkets & Grocery Stores	1,673		1,793	7.0	-0.3	27.0
Takeaway Food	374		271	-27.5	-2.2	6.0
Liquor	282		309	9.9	0.9	4.5
Other Food	202		215	6.4	0.8	3.3
Hospitality & Service Industries						
Hotels & Licensed Clubs	647		747	15.5	0.9	10.4
Cafes & Restaurants	456		389	-14.8	-0.7	7.4
Selected Services	56		59	5.4	-4.8	0.9
Other, Pharma, Watches						
Other Retailing	243		256	5.3	-1.9	3.9
Pharmaceutical, Cosmetic & Toiletry	216		266	23.3	0.2	3.5
Watch & Jewellery	93		103	10.6	-0.4	1.5
Department Stores						
Department Stores	482		500	3.8	0.1	7.8
Household Goods						
Furniture & Floor Covering	407		464	14.0	1.6	6.6
Domestic Hardware & Houseware	282		254	-10.0	-0.2	4.6
Domestic Appliances & Recorded Music	169		214	27.0	-1.0	2.7
Clothing & Soft Goods						
Clothing	308		331	7.6	0.4	5.0
Other Clothing Related	113		122	7.9	0.0	1.8
Recreational Goods						
Newspaper, Book & Stationery	113		139	22.7	1.2	1.9
Other Recreational Goods	77		83	7.4	-1.0	1.2

نموذج 5–5. نصائح توفتي: Sparklines (غريغور آيش)

بشكل عام، تتيح لك الرسوم البيانية وضع تصور أو تخطيط الأبعاد في بياناتك لإيضاح الخصائص المرئية للأشكال الهندسية. وقد كُتب الكثير عن فعالية الخصائص المرئية الفردية، وخلاصة ما كُتب: أن الألوان معقدة والموقع يمثل كل شيء. ففي مخطط التشتت scatterplot على سبيل المثال، يتم عرض بُعدين في الموضعين X وY. ويمكنك كذلك أن تعرض بُعداً ثالثاً بتغيير بلون أو حجم الرموز المعروضة. أما الرسوم البيانية الخطية فهي مناسبة خصيصاً لتوضيح التطور الزمني، بينما الرسوم البيانية الشريطية مثالية للمقارنة بين البيانات المصنفة إلى

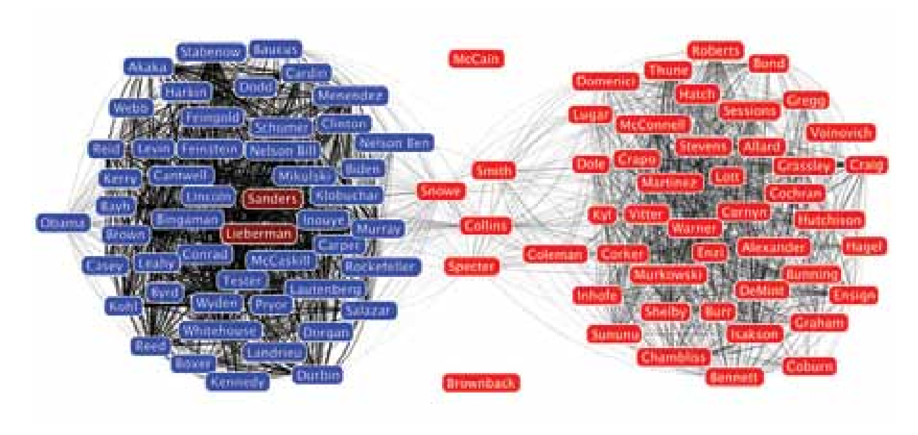
فئات. يمكنك أن تراكم عناصر الرسم البياني فوق بعضها البعض. وإذا أردت أن تقارن عدداً صغيراً من المجموعات في بياناتك، فإن عرض حالات متعددة على نفس الرسم البياني يعتبر طريقة قوية للغاية (ويشار إليها كذلك باسم المُضاعِفات الصغيرة small multiples). يمكنك في كل الرسوم البيانية أن تستخدم مختلف أنواع المقاييس لاستكشاف جوانب مختلفة في بياناتك (كالمقياس الخطي أو اللوغاريتمي).

وفي الواقع، فإن معظم البيانات التي نتعامل معها بشكل ما، تتعلق بأناس حقيقيين. وتكمن قوة الخرائط في إعادة توصيل البيانات بعالمنا المادي. تخيل مجموعة بيانات للمواقع الجغرافية للجرائم، حيث ترغب أساساً في أن ترى مكان وقوع الجرائم. كما تستطيع الخرائط أن تظهر العلاقات الجغرافية داخل البيانات، مثل التوجه من الشمال إلى الجنوب أو من المدن إلى الريف.



نموذج 5–6. خرائط التظليل النسبي (كوروبلِث) Choropleth map (غريغور آيش)

وبمناسبة الحديث عن العلاقات، فإن رابع أهم الأنواع من التوضيح المرئي هو الرسم البياني (graph) والهدف منه إظهار الصلات الداخلية (الحواف) في نقاط البيانات المتقاطعة (العقد). ويتم حساب موقع العقد عبر خوارزميات معقدة نسبيا لتخطيط الرسم البياني تسمح لنا برؤية فورية لهيكل الشبكة. وعموماً تكمن حيلة التوضيح المرئي للرسم البياني في العثور على طريقة مناسبة لتشكيل الشبكة نفسها، فليست كل مجموعات البيانات تشمل بالفعل صلات وحتى وإن كانت تشملها فقد لا تكون أكثر الجوانب إثارة للاهتمام. وأحياناً يعود القرار إلى الصحفي في تحديد الحواف بين العقد. ومن الأمثلة المثالية على ذلك ’الرسم البياني الاجتماعي لمجلس الشيوخ الأمريكي‘ (http://slate.me/senate–social)، والذي تصل حوافه بين أعضاء مجلس الشيوخ الذين أدلوا جميعا بنفس الأصوات فيما يزيد عن 65٪ من الاقتراعات بالتصويت.



نموذج 5–7. توضيحات مرئية لرصد الأنماط المسببة (غريغور آيش)

حلل وفسر ما تراه

بمجرد توضيحك المرئي لبياناتك، ستكون الخطوة التالية التعرف على أمور ما من الصورة التي وضعتها. يمكنك أن تسأل نفسك:

- ما الذي يمكنني أن أراه في هذه الصورة؟ هل هي ما توقعته؟
- هل هناك أية أنماط مثيرة للاهتمام؟
- ماذا يعني ذلك في سياق البيانات؟

قد ينتهي بك الأمر أحياناً لوضع توضيح مرئي قد يبدو رغم جماله أنه لا يخبرك بأي شيء مهم عن بياناتك. ولكن دائماً تقريباً ما يوجد شيء ما يمكنك معرفته من أي توضيح مرئي، مهما كان تافهاً.

وثّق أفكارك وخطواتك

إذا اعتبرت تلك العملية بمثابة رحلة عبر مجموعة البيانات، فإن التوثيق هو مذكرات سفرك، وسيخبرك بوجهة سفرك وما رأيته هناك وكيف قررت اتخاذ خطواتك التالية. يمكنك حتى أن تبدأ التوثيق قبل أن تلقي النظرة الأولى على البيانات.

في أغلب الحالات عندما نبدأ العمل مع مجموعة بيانات لم نرها من قبل، تكون لدينا بالفعل توقعات وافتراضات عديدة تتعلق بمحتواها. وعادة ما يكون هناك سبب لاهتمامنا بمجموعة البيانات التي نتعامل معها. وهي فكرة جيدة أن تبدأ التوثيق بكتابة تلك الأفكار المبدئية. وهذا يساعدنا على تحديد انحيازنا ويقلل من خطر سوء تفسير البيانات وذلك من خلال العثور على ما أردنا العثور عليه أصلاً.

أنا أعتقد حقاً أن التوثيق هو أهم خطوات العملية؛ وهو أيضاً الخطوة التي نميل إلى تجاوزها. ومثلما ترى في المثال أدناه، تتضمن العملية المشروحة الكثير من التخطيط ومعالجة البيانات. وقد يكون النظر إلى مجموعة من خمسة عشر رسماً بيانياً مربكاً للغاية، وخاصة بعد مرور بعض الوقت. وفي الحقيقة، هذه الرسوم البيانية تكتسب قيمة فقط (بالنسبة لك أو أي شخص تود أن تُطلعه على نتائجك) إذا تم تقديمها في السياق الذي وُضعت من أجله. ولذلك، فعليك أن تخصص الوقت لكتابة بعض الملاحظات حول أمور مثل:

- لماذا أعددت هذا الرسم البياني؟
- ماذا فعلت بالبيانات لأتم وضعه بهذه الصورة؟
- بماذا يخبرني هذا الرسم البياني؟

حوّل البيانات

بطبيعة الحال، مع تزايد الأفكار التي جمعتها من التوضيح المرئي الأخير، قد يتكون لديك إدراك ما حول ما تود رؤيته بعد ذلك. ربما تكون قد وجدت نمطاً

مثيراً للاهتمام في مجموعة البيانات التي تريد الآن أن تتفحصها بشكل أكثر تفصيلاً.

من التحوّلات المحتملة:

تكبير الصورة

وهو ما يعني إلقاء نظرة متمعنة على تفصيل معيّن في التوضيح المرئي وفحصها بغض تجميع العديد من نقاط البيانات المتصلة أو دمجها.

التصفية

إزالة نقاط البيانات (مؤقتا) التي تقع خارج إطار بؤرة اهتمامنا الأساسي.

إزالة القيم المتطرفة

التخلص من النقاط المنفردة التي لا تمثل 99٪ من مجموعة البيانات.

بافتراض أنك وضعت توضيحاً مرئياً لرسم بياني وما نتج عن ذلك كان مجرد فوضى من العقد المتصلة من خلال مئات الحواف (وهي نتيجة شائعة جداً عند عمل إيضاحات مرئية لما يُدعى الشبكات المتصلة بكثافة densely connected networks)، فمن الخطوات الشائعة للتحوّل تصفية بعض الحواف. إذا كانت الحواف مثلاً تمثل تدفق الأموال من الدول المانحة إلى الدول المتلقية، يمكننا إزالة أي تدفق أقل من مبلغ معين.

أي الأدوات نستخدم

إن مسألة الأدوات ليست بسيطة، فكل أداة متاحة لتوضيح البيانات مرئياً تستخدم جيدا لأمر معين. ومن البديهي أنه يجب أن يكون التوضيح المرئي ومعالجة البيانات أمرين يسهل التعامل معهما وثمنهما زهيد. وإذا كان تغيير معطيات التوضيح المرئي يستغرق منك ساعات، إن هذا يعني أنك لن تتمكن من إجراء العديد من التجارب للتعلم منها. وهذا لا يعني بالضرورة أنك لن تحتاج إلى تعلم كيفية استخدام الأداة، ولكن بمجرد أن تتعلم كيف تستخدمها، ستكون فعالة للغاية.

ومن المنطقي عادة أن تختار أداة تغطي مشاكل كلٍ من معالجة البيانات

وتوضيحها مرئياً. وضرورة فصل المهام بين الأدوات المختلفة قد يعني أنه يجب عليك استيراد وتصدير بياناتك على نحو متكرر. فيما يلي قائمة قصيرة لبعض أدوات معالجة البيانات وتوضيحها مرئياً:

- الجداول البيانية مثل LibreOffice أو إكسل أو مستندات جوجل.
- أطُر البرمجة الإحصائية مثل R (r–project.org) أو Pandas (pandas.pydata.org).
- نُظُم المعلومات الجغرافية (GIS) مثل Quantum GIS أو ArcGIS أو GRASS.
- مكتبات التوضيح المرئي مثل d3.js (mbostock.github.com/d3)، أو Prefuse (prefuse.org)) أو Flare (flare.prefuse.org).
- أدوات معالجة البيانات مثل Google Refine أو Datawrangler.
- برامج التوضيح المرئي بـدون البرمجة مثل ManyEyes أو Tableau Public(tableausoftware.com/products/public).

تم وضع نماذج التوضيح المرئي المذكورة في الباب التالي باستخدام التطبيق R التي تعتبر أفضل البرامج للتوضيح المرئي (العلمي) للبيانات.

مثال: محاولة فهم بيانات تبرعات الانتخابات الأمريكية

دعونا نلقي نظرة على قاعدة بيانات تمويل حملات انتخابات الرئاسة الأمريكية التي تحتوي على حوالي 450,000 تبرع للمرشحين الرئاسيين الأمريكيين. يبلغ حجم ملف CSV نحو 60 ميجابايت وهو أكبر كثيراً مما يمكن التعامل معه بسهولة في برنامج مثل إكسل.

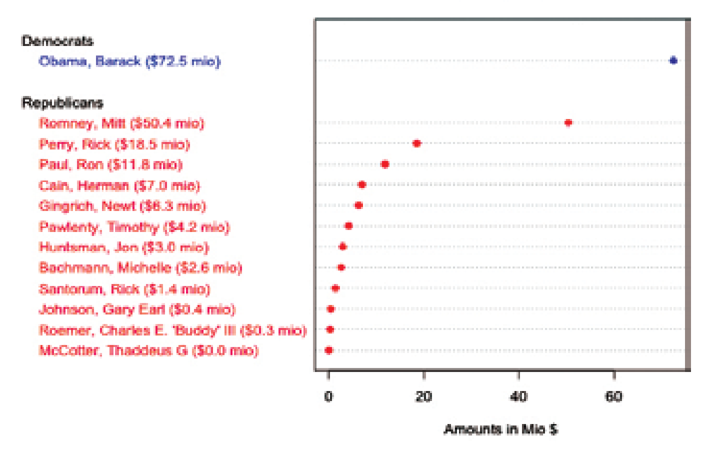
في الخطوة الأولى سأكتب بشكل صريح افتراضاتي المبدئية بخصوص مجموعة بيانات تبرعات اللجنة الفيدرالية للانتخابات:

- يحصل أوباما على أغلب التبرعات (بصفته الرئيس والمرشح الأكثر شعبية)
- يتزايد عدد التبرعات مع اقتراب موعد الانتخابات
- يحصل أوباما على الكثير من التبرعات الصغيرة، أكثر مما يحصل عليه المرشحون الجمهوريون

للإجابة على السؤال الأول علينا أن نقوم بتحويل البيانات. وبدلاً من كل تبرع منفرد، نحتاج إلى جمع إجمالي المبالغ التي تم التبرع بها لكل مرشح. وبعد توضيح النتائج مرئياً في جدول منظم، يمكننا تأكيد افتراضنا بأن أوباما يحصل على الجزء الأكبر من الأموال:

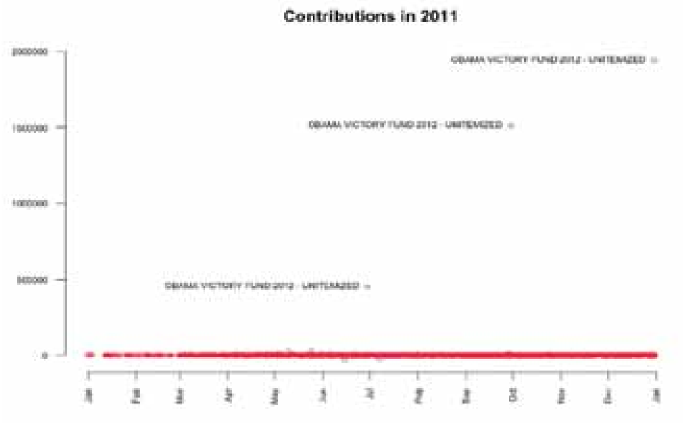
Candidate	($) Amount
Obama, Barack	72,453,620.39
Romney, Mitt	50,372,334.87
Perry, Rick	18,529,490.47
Paul, Ron	11,844,361.96
Cain, Herman	7,010,445.99
Gingrich, Newt	6,311,193.03
Pawlenty, Timothy	4,202,769.03
Huntsman, Jon	2,955,726.98
Bachmann, Michelle	2,607,916.06
Santorum, Rick	1,413,552.45
Johnson, Gary Earl	413,276.89
Roemer, Charles E. Buddy III	291,218.80
McCotter, Thaddeus G	37,030.00

وعلى الرغم من أن هذا الجدول يظهر الحد الأدنى والحد الأقصى للمبالغ وترتيبها، لا يخبرنا بالكثير عن الأنماط المتعلقة بترتيب المرشح. ويمكننا إلقاء نظرة أخرى على البيانات من خلال نموذج 5–7 وهو نوع من الرسوم البيانية يُدعى "رسم بياني بالنقط" (dot chart) يتيح لنا رؤية كل ما رأيناه في الجدول إضافة إلى الأنماط داخل المجال. فمثلاً، يسمح لنا الرسم البياني بالنقط بأن نقارن فوراً المسافة بين أوباما ورومني وبين رومني وبيري بدون الحاجة إلى طرح القيم. (ملاحظة: تم وضع الرسم البياني بالنقط باستخدام R. يمكنك أن تجد روابط للكود المصدري في نهاية هذا الفصل).



نموذج 5–8. ثلاث قيم متطرفة واضحة (غريغور آيش)

دعونا نواصل الآن بصورة أكبر لمجموعة البيانات. كخطوة أولى وضعت توضيحاً مرئياً لكل مبالغ التبرعات عبر فترة زمنية في مخطط بسيط. نستطيع رؤية أن كل التبرعات صغيرة للغاية مقارنةً بثلاث قيم متطرفة جداً. وقد أوضح المزيد من التحقق أن تلك التبرعات الضخمة تأتي من "صندوق انتصار أوباما 2012" (Obama Victory Fund 2012) (والذي يعرف أيضاً باسم 'اللجنة السياسية الخارقة' Super PAC) وقد تم منحها (450 ألف دولار) في 29 يونيو، و(1.5 مليون دولار) في 29 سبتمبر، و(1.9 مليون دولار) في 30 ديسمبر.



نموذج 5–9. إزالة القيم المتطرفة (غريغور آيش)

وبينما تعتبر تبرعات اللجان السياسية الخارقة في حد ذاتها أكبر خبر في البيانات بلا شك، قد يكون من المثير للاهتمام أيضاً أن ننظر إلى أبعد من ذلك. المقصود الآن هو أن تلك التبرعات الكبيرة تربك آراءنا حول التبرعات الأصغر الآتية من الأفراد، لذا سنقوم بإزالتها من البيانات. يُدعى هذا التحوُل إزالة القيم المتطرفة. وبعد التوضيح المرئي مرة أخرى، يمكننا أن نرى أن معظم التبرعات تتراوح بين خمسة آلاف وعشرة آلاف دولار.

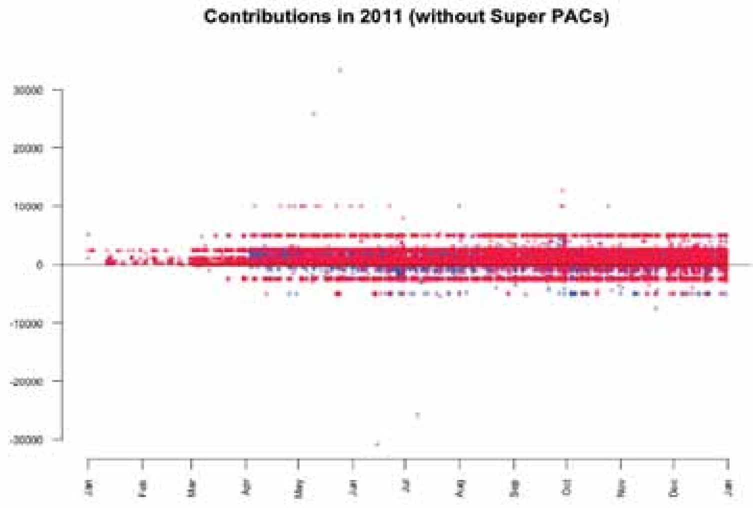
وفقاً لحدود التبرعات التي وضعها قانون اللجان الفيدرالية للانتخابات، لا يُسمح للأفراد بالتبرع بأكثر من 2500 دولار لكل مرشح. وكما نرى في المخطط، هناك العديد من التبرعات التي تتجاوز هذا الحد الأقصى. وما يجذب انتباهنا تبرعان كبيران على وجه الخصوص. ويبدو أنهما منعكسان في مبالغ سلبية (مبالغ مستردة) في شهري يونيو ويوليو. وكشف المزيد من التحقق في البيانات عن المعاملات المالية التالية:

- في 10 مايو تبرع ستيفين جيمس دايفيس من سان فرانسيسكو والمحامي في شركة Banneker Partners بمبلغ 25,800 دولار لأوباما.
- في 25 مايو تبرعت سينثيا ميرفي من ليتل روك والتي تعمل في قسم العلاقات العامة في Murphy Group بمبلغ 33,300 دولار لأوباما.
- في 15 يونيو تم رد مبلغ 30,800 دولار لسينثيا ميرفي مما خفض قيمة التبرع إلى 2500 دولار.
- في 8 يوليو تم رد مبلغ 25,800 دولار لستيفين جيمس دايفيس مما خفض قيمة التبرع إلى صفر.

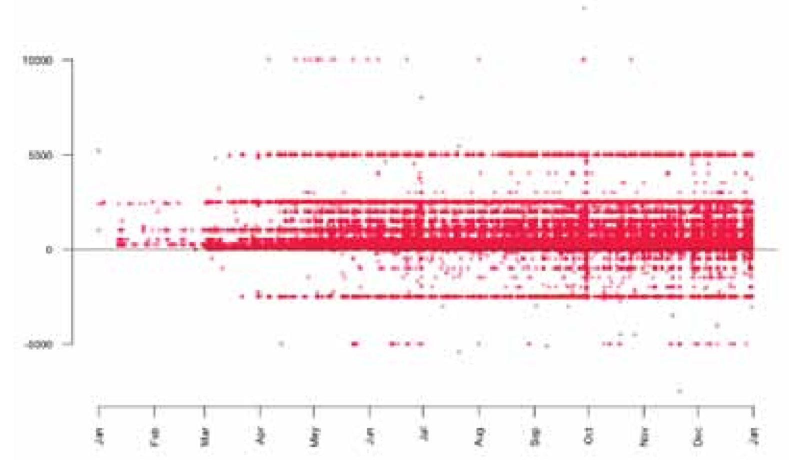
ما الذي يثير الاهتمام بخصوص تلك الأرقام؟ إن مبلغ 30,800 دولار الذي تم رده لسينثيا ميرفي يساوي الحد الأقصى للمبلغ الذي يُسمح للأفراد بالتبرع به سنوياً للجان الأحزاب السياسية الوطنية. ربما أرادت أن تجمع كلا التبرعين في معاملة مالية واحدة والتي تم رفضها. أما مبلغ 25,800 دولار الذي تم رده لستيفين جيمس دايفيس فمن المحتمل أن يساوي مبلغ 30,800 دولار ناقص 5000 دولار (وهو الحد الأقصى للتبرع لأي لجنة سياسية أخرى).

ومن النتائج الأخرى المثيرة للاهتمام في المخطط الأخير النمط الخطي

للتبرعات الممنوحة إلى المرشحين الجمهوريين بمبالغ 5000 دولار وسالب 2500 دولار. ولأتمكن من رؤيتها بمزيد من التفاصيل، وضعت توضيحاً مرئياً لتبرعات الجمهوريين فقط. والرسم البياني الناتج عن ذلك هو نموذج مثالي لأنماط البيانات التي تكون محجوبة بدون توضيح البيانات مرئياً.

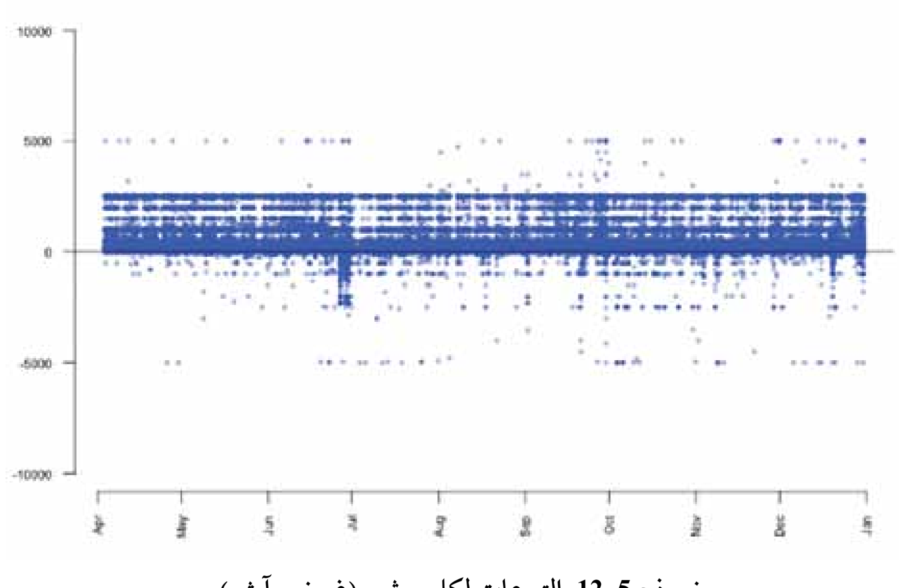


نموذج 5–10. إزالة القيم المتطرفة 2 (غريغور آيش)



نموذج 5–11. إزالة القيم المتطرفة 3 (غريغور آيش)

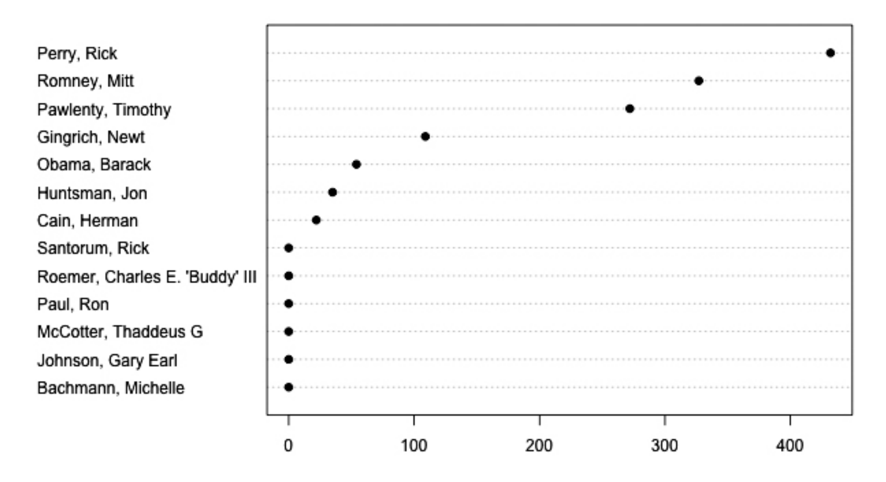
ما يمكن أن نراه هنا هو أن هناك العديد من التبرعات بمبالغ 5000 دولار للمرشحين الجمهوريون. وفي الحقيقة، يظهر لنا البحث في البيانات أن هناك 1243 تبرعا، وهي تمثل نسبة 0.3٪ من مجموع التبرعات، ولكن نظرا لأن هذه التبرعات موزعة بالتساوي عبر فترة زمنية، يظهر الخط بهذا الشكل. وما يثير الانتباه بخصوص هذا الخط هو أن الحد الأقصى لتبرعات الأفراد كان 2500 دولار. وبالتالي تم رد كل دولار تجاوز هذا المبلغ إلى المتبرعين، مما أدى إلى ظهور النمط الخطي الثاني بسالب 2500 دولار. وعلى العكس من ذلك، لا تظهر تبرعات باراك أوباما نمطاً مشابهاً.



نموذج 5–12. التبرعات لكل مرشح (غريغور آيش)

إذن قد يكون مثيراً للاهتمام أن نعرف لماذا لم يلاحظ مئات المتبرعين الجمهوريين الحد الأقصى لتبرعات الأفراد. ولتحليل هذا الموضوع بشكل أكبر، يمكننا أن نلقي نظرة على مجموع التبرعات بمبلغ 5000 دولار لكل مرشح.

هذه طبعاً نظرة مشوهة لأنها لا تنظر في إجمالي التبرعات التي تلقاها كل مرشح. يظهر المخطط التالي نسبة التبرعات بمبلغ 5000 دولار لكل مرشح.



نموذج 5–13. من أين تأتي أموال عضو مجلس الشيوخ؟
التبرعات لكل مرشح (غريغور آيش)

الدرس المستفاد

عادة ما يبدو مثل هذا التحليل المرئي لمجموعة بيانات جديدة كأنه رحلة مثيرة إلى بلد مجهول. فأنت تبدأ كأجنبي ولديك فقط البيانات ومعها افتراضاتك، ولكن مع كل خطوة تخطوها وكل رسم بياني تضعه، تجد أفكاراً جديدة حول الموضوع. وبناء عليها، تتخذ قرارات بخصوص خطواتك التالية والقضايا الجديرة بالمزيد من التدقيق. وكما رأيت في هذا الفصل، يمكن تكرار هذه العملية من التوضيح المرئي والتحليل والتحويل للبيانات بشكل لانهائي تقريباً.

احصل على الترميز 'الكود' المصدري

تم وضع جميع الرسوم البيانية المذكورة في هذا الفصل باستخدام برنامج R المتميز والذي صنع أساساً كأداة علمية للتوضيح المرئي. ومن الصعب أن تجد أي تقنية توضيح مرئي أو معالجة بيانات ليست موجودة بالفعل في برنامج R. فيما يتعلق بهؤلاء المهتمين بكيفية معالجة البيانات وتوضيحها مرئياً باستخدام R، ها هو الكود المصدري للرسوم البيانية المذكورة في هذا الفصل:

- المخطط المنقط dotchart: التبرعات لكل مرشح (http://gist.github.com/1769733).
- المخطط: جميع التبرعات عبر فترة زمنية (https://gist.github.com/1816161).
- المخطط: تبرعات اللجان المعتمدة (https://gist.github.com/1816169).

هناك أيضاً مجموعة كبيرة من الكتب والدروس التعليمية المتاحة.

— غريغور آيش، مؤسسة المعرفة المفتوحة

الفَصْل السَّادس

تقديم البيانات



عندما تلقي نظرة فاحصة على البيانات التي لديك، ثم تقرر أن ثمة أمراً مهماً تريد أن تكتب عنه، يبرز السؤال: كيف يمكنك تقديمه إلى الناس؟ يستعرض هذا الفصل طرفاً من تجارب عدد من أبرز صحافيي البيانات تحكي كيف قاموا بتقديم المعلومات التي لديهم إلى القراء، بدءاً من الرسوم البيانية ووصولاً إلى منصات البيانات المفتوحة وروابط التحميل. ثم نتناول بشيء من التفصيل طريقة إنشاء التطبيقات الإخبارية والتعرف على أسرار التمثيل المرئي للبيانات. وأخيراً، سنلقي نظرة على ما يمكنك القيام به لتفعيل مشاركة الجمهور في المشروع الخاص بك.

تقديم البيانات للجمهور

توجد طرق عديدة لتقديم البيانات التي لديك للجمهور، منها نشر مجموعات بيانات أولية مع الأخبار، وإضافة توضيحات مرئية جميلة، وإنشاء تطبيقات إنترنت تفاعلية. وقد سألنا بعضاً من كبار صحافيي البيانات عن نصائحهم حول طريقة تقديم البيانات للجمهور.

هل نستخدم التوضيحات المرئية أم لا نستخدمها؟

أحياناً تستطيع البيانات أن تسرد الخبر بشكل أفضل مما تفعل الكلمات أو الصور، ولهذا ظلت المصطلحات مثل "تطبيق الأخبار" و"التوضيح المرئي للبيانات" تحظى بصدى واسع في العديد من غرف الأخبار مؤخرًا. وقد تزايد هذا الاهتمام بسبب النتاج الهائل من الأدوات والتقنيات الجديدة (المجانية غالباً) التي صُممت لتمكن الصحافيين– حتى الذين يجدون صعوبةً في التعامل مع التكنولوجيا– من تحويل البيانات إلى سرد مرئي للأخبار.

فهناك أدوات مثل «جداول جوجل فيوجن» (Google Fusion Tables) و"ميني آيز" (Many Eyes) و"تابلو" (Tableau) و"ديبيتي" (Dipity) وغيرها ساعدت بشكل كبير في إنشاء الخرائط والمخططات والرسوم البيانية، إضافة إلى تطبيقات البيانات المتكاملة التي ما زالت نطاق عمل المتخصصين حتى الآن. ومع زوال الحاجز أمام دخول المجال، يظل السؤال الذي يواجه الصحافيين ليس في مدى قدرتهم على تحويل مجموعة البيانات إلى توضيح مرئي، بل ما إذا كان من الضروري عليهم فعل ذلك. فالتوضيح المرئي السيء (http://bit.ly/niemanlab-wordcloud) أسوأ كثيراً من عدمه.

— آرون بيلهوفر، صحيفة نيويورك تايمز

استخدام الرسوم البيانية المتحركة

مع وجود نص مُحكم ورسوم متحركة ذات توقيت مناسب وشرح واضح، تستطيع الرسوم البيانية المتحركة تقريب الأرقام أو الأفكار المعقدة للأذهان والانتقال بالمشاهد خلال الخبر بسهولة وسلاسة. وتعد محاضرات الفيديو التي يلقيها هانز روزلينغ أحد الأمثلة الجيدة على كيفية بعث الحياة في البيانات لتخبرك بشيء ما على الشاشة. وسواء أكنت متفقاً مع منهجهم أم مختلفاً، أعتقد أيضاً أن 'مؤشر قاذف الحذاء' في مجلة الإكونوميست (http://econ.st/shoethrow-ers) يُعتبر مثالاً جيداً على استخدام الفيديو لسرد خبر قائم على الأرقام. لا تقم بعرض هذا الرسم البياني في صورة ثابتة وتحاش ذلك قدر الإمكان. فالأمور تتطور بسرعة، ولكن مع الإعداد لذلك خطوة بخطوة، ستفهم كيف ولماذا وصلوا إلى هذا المؤشر. ويمكنك من خلال الرسوم البيانية المتحركة ومقاطع الفيديو المتحركة القصيرة تعزيز التسجيل الصوتي الذي يسمعه جمهورك إلى جانب الشرح المرئي، مما يوفر وسيلة فعالة للغاية لسرد الخبر يصعب نسيانها.

—— لولو بيني، بي. بي. سي.

إبلاغ العالم

يبدأ سير العمل لدينا عادة ببرنامج إكسل، لأنها طريقة سهلة للغاية تمكننا من معرفة ما إذا كان هناك ما يثير الاهتمام في البيانات بسرعة. فإذا شعرنا بوجود مادة مثيرة في البيانات، نذهب إلى مكتب الأخبار، ونحن محظوظون جداً لموقعنا المجاور لمكتب الأخبار الرئيسي في صحيفة الغارديان. ثم نبحث في كيفية تصوير تلك البيانات أو وضعها في صورة عرض مرئي على الصفحة، وبعد ذلك نكتب ما يُنشر معها. وعندما أكتب، عادة ما تكون معي نسخة مختصرة من الجدول البياني أضعها إلى جانب محرر النص. غالباً ما أقوم ببعض التحليلات أثناء الكتابة لأنتقي الأمور المثيرة للاهتمام. ثم أنشر المقالة وأقضي بعض الوقت في التغريد حولها، والكتابة إلى عدة أشخاص، وضمان وجود روابط لها من جميع الأماكن المناسبة.

تصل نصف الزيارات لمقالاتنا من خلال موقعي تويتر وفيسبوك، ونحن نفخر بأن متوسط الوقت الذي يقضيه القارئ في الاطلاع على مقالة على مدونة

البيانات هو 6 دقائق، مقارنة بمتوسط يبلغ دقيقة واحدة يقضيها على بقية موقع الغارديان. فست دقائق تُعتبر رقماً جيداً جداً، حيث إن الوقت الذي يقضيه القارئ على الصفحة يُعتبر من أهم المقاييس عند تحليل زيارات التصفح.

وهذا يساعد أيضاً في إقناع زملائنا بقيمة عملنا، فبالإضافة إلى الأخبار الهائلة المبنية على البيانات التي عملنا عليها ويعرفها الجميع في غرفة الأخبار، هناك قاعدة بيانات COINS وويكيليكس وأحداث الشغب في المملكة المتحدة. في قاعدة بيانات COINS الخاصة بالمصروفات الحكومية، كان لدينا 5–6 صحافيين متخصصين في الغارديان يعملون للإدلاء بآرائهم حول البيانات حين نشرتها الحكومة البريطانية. كما كان لدينا فريق آخر مكون من 5–6 أشخاص عندما نُشرت بيانات عن بنود الإنفاق الحكومية التي يزيد أي منها عن 25 ألف جنيه استرليني – تضمنت تلك الفرق صحافيين معروفين مثل بولي كيرتس. وبالطبع كان ويكيليكس مشروعاً ضخماً يتضمن الكثير من الأخبار عن العراق وأفغانستان. كما كان مشروع أحداث الشغب كبيراً حيث سجّل ما يقرب من 550 ألف زيارة خلال يومين.

إلا أن الأمر لا يتعلق بالنجاحات قصيرة الأمد فحسب، بل بكونه أيضاً مصدراً موثوقاً ومفيداً للمعلومات، ونحن نسعى إلى أن نكون أحد المواقع التي يمكنك أن تحصل منها على معلومات جيدة ومفيدة حول الموضوعات التي نغطيها.

—— سايمون روجرز، صحيفة الغارديان

نشر البيانات

عادة ما ندرج البيانات التي لدينا على موقعنا الإلكتروني في توضيح مرئي وبشكل يتيح تحميل مجموعة البيانات بسهولة، يستطيع قراؤنا أن يستكشفوا البيانات المضمنة في الأخبار من خلال التفاعل مع التوضيح المرئي أو استخدام البيانات بأنفسهم بطرق مختلفة. وتبرز أهمية ذلك في تعزيز الشفافية لدى صحيفة سياتل تايمز، من خلال عرض نفس البيانات التي استخدمناها في الوصول إلى استنتاجات مفيدة وإتاحتها للقراء. الذين يستخدمون هذه البيانات هم نقّادنا بالطبع، بالإضافة إلى شريحة القراء التي تهتم فقط بالخبر وما يتعلق به ويترتب عليه من نتائج. ويمكننا أيضاً من خلال إتاحة البيانات وضع قائمة بالملاحظات التي أبداها

هؤلاء النقاد وعموم القراء حول ما قد يفوتنا وما يمكننا أن نستكشفه – وكل ذلك مفيد في سبيل الوصول إلى الصحافة الجادة.

—— شيريل فيليبس، صحيفة سياتل تايمز

إتاحة البيانات

إن إتاحة الفرصة أمام العملاء الجدد للحصول بسهولة على البيانات التي نستخدمها في عملنا هو الصواب، وذلك لعدة أسباب؛ منها اطمئنان القراء إلى أننا لا نلوي عنق البيانات لنتوصّل إلى استنتاجات غير منصفة. وتعتبر إتاحة البيانات التي لديك أحد الأمور المتعارف عليها في العلوم الاجتماعية التي تسمح للباحثين بتكرار الأعمال التي قمت بها. تشجيع القراء على دراسة البيانات من شأنه أن يولِّد أفكاراً قد تفضي إلى أخبار متلاحقة. وفي النهاية، من المرجح أن يعود القراء المشاركون المهتمون بهذه البيانات مرة تلو الأخرى.

—— ستيف دويغ، كلية والتر كرونكايت للصحافة، جامعة أريزونا

إنشاء منصة البيانات المفتوحة

في صحيفة لا ناسيون، يُعتبر نشر البيانات المفتوحة جزءاً أساسياً من أنشطة صحافة البيانات لدينا. لا يوجد في الأرجنتين قانون لحرية تبادل المعلومات أو بوابة بيانات وطنية، لهذا نهتم كثيراً بتزويد قرائنا بالبيانات التي نستخدمها في أخبارنا.

لذا ننشر البيانات المنظمة المجردة من خلال منصة Junar المتكاملة، بالإضافة إلى جداول جوجل البيانية. كما نسمح ونشجع الآخرين صراحة على إعادة استخدام بياناتنا ونشرح قليلاً كيفية عمل ذلك بالتوثيق ومقاطع الفيديو التعليمية (http://bit.ly/lanacion-tutorials).

علاوة على ذلك، فإننا نقدم بعضاً من مجموعات البيانات والتوضيحات المرئية تلك في مدونة بيانات لا ناسيون (http://blogs.lanacion.com.ar/data/)، بغرض التوعية حول البيانات وأدوات نشرها في الأرجنتين، ولنوضح للآخرين كيف قمنا بجمع البيانات التي لدينا وكيف نستخدمها وكيف يمكنهم إعادة استخدامها.

منذ افتتاحنا هذه المنصة في فبراير 2012، تلقينا اقتراحاتٍ وأفكاراً لمجموعات البيانات، كان أغلبها من الأكاديميين والباحثين، بالإضافة إلى طلاب الجامعات الذين يعربون عن الامتنان لنا كثيراً في كل مرة نرد عليهم بأحد الحلول أو بمجموعة بيانات محددة. كما يتفاعل القراء مع بياناتنا في تابلو Tableau ويعلقون عليها. وقد حصلنا في مرات عديدة على أعلى نسبة تعليقات ومشاهدات بين من يقدمون هذه الخدمة. وقد كان لدينا 7 من بين أكثر 100 توضيح مرئي حققت أعلى نسبة مشاهدة في عام 2011 (http://bit.ly/tableau-7-100).

— أنجيليكا بيرلاتا راموس، صحيفة لا ناسيون (الأرجنتين)

دمج العنصر البشري في البيانات

مع اتساع النقاش حول حدود البيانات الضخمة ليصل إلى الوعي العام الأكبر، كان هناك جزء هام مفقود بوضوح، وهو العنصر البشري. يرى البعض أن البيانات عبارة عن أرقام منفصلة وغير مترابطة، إلا أنها في الحقيقة مقاييس لأمور ملموسة (بشرية في أكثر الأحيان). فالبيانات ترتبط بالحياة الحقيقية لأشخاص حقيقيين، وعندما نتعامل مع الأرقام، علينا أن نأخذ في الحسبان أنظمة العالم الحقيقي التي أتت منها.

خذ بيانات تحديد الموقع، على سبيل المثال، التي يتم جمعها الآن من مئات الملايين من الهواتف والأجهزة المحمولة. من السهل الاعتقاد بأن تلك البيانات (الأرقام التي تمثل خطوط العرض والطول والوقت) هي 'عوادم رقمية'، ولكنها في الواقع لحظات منتقاة من حكاياتنا الشخصية التي قد تبدو جافة ومجردة عند قراءتها في جدول بياني. ولكن عندما نسمح للناس بوضع بياناتهم الخاصة على خريطة وإعادة عرضها، تجدهم يعيشون حالة استعادة للذاكرة تتسم بالقوة والإنسانية.

تُستخدم بيانات تحديد الموقع حالياً من قبل العديد من 'الأطراف الأخرى' - مثل مطوري التطبيقات والأسماء التجارية الكبيرة وأصحاب الإعلانات. وفي الوقت الذي تملك فيه 'الأطراف الثانية' (شركات الاتصالات ومدراء الأجهزة) هذه البيانات، نجد أن 'الطرف الأول' - وهو أنت - لا يملك الحصول أو

السيطرة على تلك المعلومات. ففي قسم البحوث والتطوير بصحيفة نيويورك تايمز، أطلقنا مشروعاً نموذجياً وأسميناه OpenPaths (https://openpaths.cc) لنسمح للناس باستكشاف بيانات مواقعهم الخاصة وإدراك مفهوم ملكية البيانات. وفي النهاية، يجب أن يتمكن الناس من التحكم في تلك الأرقام المرتبطة بحياتهم وتجاربهم بشكل كبير.

للصحافيين دور بالغ الأهمية في إلقاء الضوء على هذا العنصر البشري الكامن في البيانات. وبذلك، تكون لديهم القدرة على تغيير المفهوم العام لكل من البيانات والأنظمة التي تأتي منها الأرقام.

— جير ثورب، برنامج Data Artist in Residence: قسم البحوث والتطوير بصحيفة نيويورك تايمز

البيانات المفتوحة، المصادر المفتوحة، الأخبار المفتوحة

يمكننا أن نعتبر عام 2012 عام الأخبار المفتوحة. فهي في صميم العقيدة التحريرية لدينا وتُعتبر رسالة هامة للاسم التجاري الحالي الخاص بنا. وفي خضم كل ذلك، بات من الواضح أننا بحاجة إلى عملية مفتوحة لصحافة البيانات، على ألا يكون محرك هذه العملية البيانات المفتوحة فحسب، بل لا بد من تفعيل الأدوات المفتوحة أيضاً. نأمل بحلول نهاية العام أن نتمكن من تزويد كل توضيح مرئي ننشره بإمكانية الوصول إلى كل من البيانات المستخدمة فيه ورمز التفعيل الخاص به.

العديد من الأدوات المستخدمة في التوضيح المرئي اليوم مغلقة المصدر. وهناك أدوات أخرى تأتي مع تراخيص مقيدة تحظر استخدام البيانات المأخوذة منها. أما المكتبات القائمة مفتوحة المصدر فتسهم غالباً في حل مشكلة واحدة، لكنها تفشل في وضع منهجية أكثر شمولاً. وهذا كله يجعل البناء على أعمال الآخرين أمراً في غاية الصعوبة، كما أنه يسد باب المحادثات بدلاً من فتحها. ونحن نقوم بتطوير مجموعة من الأدوات المفتوحة للسرد التفاعلي للأخبار - وهو مشروع Miso (@themisoproject).

نحن بصدد مناقشة هذا العمل مع عدد من المؤسسات الإعلامية الأخرى، لأن الأمر يتطلب مشاركة جماعية لتحقيق الاستفادة الكاملة من إمكانيات برنامج

مفتوح المصدر. وإذا نجحنا، سيقدم هذا ديناميكية مختلفة تماماً لقرائنا. ويمكن أن تتجاوز إسهامات المشاركين مجرد التعليق لتصل إلى الدعم المالي للأعمال أو إصلاح المشاكل أو إعادة استخدام البيانات بطرق جديدة.

—— ألاستير دانت، صحيفة الغارديان

إضافة روابط التنزيل

في السنوات القليلة الماضية، تعاملت مع كمية ضخمة من البيانات لمشاريع أو مقالات يبلغ حجمها بضعة غيغابايتات، من مسح للجداول المطبوعة من الستينيات إلى 1.5 غيغابايت من البرقيات التي ينشرها موقع ويكيليكس. كان من الصعب دائماً أن نقنع المحررين بنشر المصادر الرئيسية للبيانات بشكل منتظم في صيغة مفتوحة يمكن للجميع الوصول إليها. ولتفادي هذه المشكلة، قمت بإضافة روابط 'تنزيل البيانات' داخل المقالات تنتقل مباشرة إلى الأرشيف الذي يحتوي على الملفات أو وثائق جوجل ذات الصلة. كان اهتمام الأشخاص المتوقع استفادتهم من الخدمة مرة أخرى متماشياً مع ما نراه في البرامج التي ترعاها الحكومة (أي ضئيلاً للغاية)؛ إلا أن تكرار الاستفادة من الخدمة – وإن كان قليلاً – زودنا بأفكار جديدة، وأثار نقاشات جديرة بالأوقات القليلة التي قضيناها في العمل على كل مشروع!

—— نيكولاس كايزر بريل، شبكة Journalism++

تحديد نطاق العمل

حدد نطاق العمل. هناك فرق كبير بين البرمجة لمجرد التسلية وبين الهندسة من أجل المقاييس والأداء. تأكد من أن المشاركين معك في هذا المشروع لديهم مجموعة المهارات المناسبة، دون إغفال لدور التصاميم. حيث تعتبر كل من صلاحية المشروع وخبرات المستخدمين وتصاميم العروض عوامل مؤثرة في نجاح مشروعك بشكل كبير.

—— كريس وو، ملتقى الصحافيين والمبرمجين Hacks/Hackers

كيفية إنشاء تطبيق الأخبار

تُعتبر تطبيقات الأخبار بمثابة نوافذ على بيانات الخبر؛ وهذه قد تكون قواعد بيانات قابلة للبحث أو توضيحات مرئية جذابة أو شيء آخر مختلف تماماً. وبغض النظر عن شكلها، تشجع تطبيقات الأخبار القراء على التفاعل مع البيانات في إطار مفيد بالنسبة لهم: كالبحث عن توجهات الجرائم في منطقتهم أو التحقق من سجلات السلامة لطبيبهم المحلي أو البحث عن الإسهامات السياسية للمرشح الذي اختاروه.

وتعتبر التطبيقات المستقرة هي الأفضل في تطبيقات الأخبار، وتتفوق أيضاً على الرسوم البيانية ذات التقنية العالية، حيث إنها تقع خارج دورة الأخبار من خلال مساعدة القراء على حل المشاكل الحياتية الحقيقية أو الإجابة على الأسئلة بطريقة مفيدة وجديدة تجعل منها موارد مستمرة. عندما أراد الصحافيون في ProPublica استكشاف مدى سلامة عيادات غسيل الكلى الأمريكية، أنشأوا تطبيقاً (http://projects.propublica.org/dialysis/) يساعد القراء على تقييم سلامة عياداتهم المحلية. لا شك في أن توفير مثل هذه الخدمة الهامة يبني علاقات مع المستخدمين تتجاوز النتائج التي يمكن تحقيقها من خلال رواية الخبر وحده.

وهنا يكمن التحدي والتعهد بإنشاء تطبيقات أخبار متطورة، بمعني إيجاد شيء تبقى قيمته على الدوام. وسواء أكنت مطوّراً أم مديراً، يجب أن يبدأ أي نقاش حول كيفية إنشاء تطبيق متميز للأخبار بعقلية تطوير المنتج: ركز مباشرة على المستخدم واعمل من أجل تحقيق الاستفادة القصوى مقابل الجهد المبذول. ومن المفيد جداً طرح ثلاثة أسئلة قبل البدء في إنشاء التطبيق:

من هم جمهوري وماهي احتياجاتهم؟

لا تقوم تطبيقات الأخبار بخدمة الخبر نفسه، بل تفيد المستخدم. وبحسب المشروع، قد يكون هذا المستخدم أحد المرضى الراغبين في عمل غسيل للكُلى ويريد أن يتعرف على سجل سلامة العيادة التي سيعالج فيها، أو قد يكون صاحب منزل لا يعلم بمخاطر الزلازل قرب منزله. وبغض النظر عن هوية المستخدم، يجب أن يبدأ أي نقاش حول إنشاء تطبيق الأخبار بالشريحة التي ستستخدمه، كأي



نموذج 6–1. متتبع عيادات غسيل الكلى (ProPublica)

منتج جيد.

من الممكن أن يخدم تطبيق واحد عدداً كبيراً من المستخدمين. فعلى سبيل المثال، هناك مشروع يسمى Curbwise (http://curbwise.com/) أنشأته صحيفة 'أوماها وورلد هيرالد' (بولاية نبراسكا) يخدم أصحاب المنازل الذين يرون أنه يتم فرض ضرائب باهظة عليهم، وذلك من خلال الاطلاع على أسعار العقارات المجاورة وضرائبها، كما يستفيد منه العاملون في مجال العقارات الراغبون في متابعة حركات المبيعات التي تمت مؤخراً. فالتطبيق يلبي حاجة خاصة تجعل المستخدمين في كل هذه الحالات يعاودون استخدامه مرة أخرى.

على سبيل المثال، يحتاج أصحاب المنازل للمساعدة في جمع المعلومات حول المنازل المجاورة ليتحققوا من وجود ضرائب باهظة مفروضة عليهم دون وجه حق. ولا يخفى أن عملية جمع تلك المعلومات معقدة ومهدِرة للوقت، وهي مشكلة يتصدى لحلها مشروع Curbwise عبر إعداد تقرير مبسط سهل الاستخدام (http://curbwise.com/how-to-protest) يحتوي على جميع المعلومات التي يحتاجونها للدفع بعدم قانونية الضرائب المفروضة عليهم أمام السلطات المحلية. يبيع مشروع Curbwise تلك التقارير مقابل 20 دولاراً، ويدفع الناس هذا المبلغ

لأن التقرير يحل مشكلة حقيقية في حياتهم.

سواء أكان التطبيق الخاص بك يعمل على حل مشاكل حقيقية، مثلما يفعل تطبيق Curbwise، أم يكمل سرد الخبر بتوضيح مرئي مثير للاهتمام؛ عليك أن تكون مدركاً دائماً لأحوال المستخدمين، ومن ثمّ التركيز على تصميم الخصائص وإنشائها حسب احتياجاتهم.

ما المدة اللازمة لإنجاز هذا التطبيق؟

يندر وجود المطوّرين في غرفة الأخبار ندرة المياه في الصحراء. وإنشاء تطبيقات الأخبار يعني تحقيق التوازن بين المتطلبات اليومية لغرفة الأخبار والالتزامات التي يحتاجها إيجاد منتجات متميزة بالفعل على المدى البعيد.

فلنفترض أن المحرر جاءك بفكرة تدور حول إجراء مجلس المدينة تصويتاً الأسبوع القادم بشأن هدم عدة منشآت تاريخية في مدينتك. يقترح المحرر إنشاء تطبيق بسيط يسمح للمستخدمين برؤية المباني على خريطة.

بصفتك مُطوِّراً، لديك بضعة خيارات: يمكنك أن تستعرض عضلاتك في الهندسة وتنشئ خريطة جيدة باستخدام برامج متخصصة، أو يمكنك أن تستخدم أدوات موجودة بالفعل مثل جداول جوجل فيوجن أو مكتبات مفتوحة المصدر للخرائط وتنتهي من المهمة في غضون ساعات. سيعطيك الخيار الأول تطبيقاً أفضل، أما الثاني فقد يمنحك المزيد من الوقت لإعداد شيء آخر مع فرصة أفضل ليكون للتطبيق تأثير دائم.

استدعاء الخبر لوجود أحد تطبيقات الأخبار الرائعة والمعقدة لا يعني بالضرورة قيامك بإنشاء هذا التطبيق، يجب أولا تحقيق التوازن بين الأولويات. وتذكّر دائماً أن كل تطبيق تبنيه ستكون له تكلفة مادية وزمنية: ويعني ذلك أنه من الممكن إنشاء تطبيق آخر أكثر تأثيراً بدلاً من الأول.

كيف يمكنني الانتقال إلى المرحلة التالية؟

قد يكون بناء تطبيقات أخبار متطورة مكلفاً ومهدراً للوقت. لذلك من المفيد دائماً أن نسأل عن النتيجة، وكيف يمكن تطويرها إلى شيء مميز؟

من وسائل تحقيق ذلك إيجاد مشروع دائم يتجاوز حدود دورة الأخبار؛ وهناك طريقة أخرى، وهي بناء أداة توفر لك الوقت لاحقا (وتوفر مصدرا مفتوحا أيضاً!)، أو القيام بإضافة تحليلات متطورة لتطبيقك لتعرف المزيد عن جمهورك.

تتجه العديد من المؤسسات نحو إنشاء خرائط إحصاء السكان لبيان التغيرات الديمغرافية في المدن التابعة لها. ولكن عندما قام فريق تطبيقات الأخبار لدى صحيفة شيكاغو تريبيون بإنشاء خريطتهم الخاصة (http://bit.ly/chicago-census)، انتقلوا إلى المرحلة التالية عبر تطوير أدوات وتقنيات لإنشاء تلك الخرائط سريعاً، ثم جعلوها متاحة (http://bit.ly/chicagotribune-maps) لتستخدمها مؤسسات أخرى.

قمنا في مركز الصحافة الاستقصائية الذي أعمل فيه بإضافة قاعدة بيانات بسيطة مزودة بخاصية البحث إلى إطار تتبعي دقيق يسمح لنا بمعرفة مدى تقدير المستخدمين لآليات الاستكشاف والاكتشاف بالمصادفة في تطبيق الأخبار الخاص بنا.

لا بأس أن تبدو مادّياً أحيانا، فكر في عائدات الاستثمار (http://bit.ly/cironline-return). توصّل إلى حل لمشكلة عامة، ابتكر طريقة جديدة لإشراك المستخدمين، اجعل بعض أجزاء عملك مفتوحة المصدر، استخدم التحليلات لتعرف المزيد عن مستخدميك، أو حاول العثور على حالات مثل Curbwise حيث يمكن لجزء من تطبيقك أن يدر عليك عائدات مالية.

الخاتمة

قطع تطوير تطبيقات الأخبار شوطاً طويلاً خلال فترة قصيرة جداً. وكانت النسخة الأولى من تطبيقات الأخبار News Apps 1.0 تماما مثل النسخة الثانية من الرسوم البيانية Infographics 2.0– أي التوضيح المرئي التفاعلي للبيانات، إلى جانب قواعد البيانات القابلة للبحث التي صُممت أساساً لتطوير سرد الأخبار. والآن يمكن تصميم العديد من تلك التطبيقات من قبل صحافيين يعملون تحت ضغوط للتسليم في مواعد نهائية باستخدام أدوات مفتوحة المصدر وجعل المطوّرين يتفرغون للتوصّل إلى أفكار أعمق.

إن النسخة الثانية من تطبيقات الأخبار News Apps 2.0 التي تتجه إليها تلك الصناعة تتلخص في الجمع بين سرد الأخبار ونقاط القوة الخاصة بالخدمات العامة للصحافة من جهة، وبين نظام تطوير المنتج وخبرات عالم التكنولوجيا من جهة أخرى. والنتيجة بلا شك ستكون طفرة كبير في ابتكار وسائل تجعل البيانات تفي بالغرض الذي صممت لأجله ومثيرة للاهتمام ومفيدة للجمهور على وجه الخصوص، وفي نفس الوقت، نأمل أن تدفع بالصحافة نحو تحقيق الأمر ذاته.

— تشيس دايفيس، مركز الصحافة الاستقصائية

تطبيقات الأخبار في بروبابليكا

تطبيق الأخبار هو عبارة عن قاعدة بيانات تفاعلية كبيرة تسرد خبراً ما؛ فيمكن اعتباره مثل أي نوع آخر من العمل الصحافي، إلا أن الفرق يكمن فقط في أنه يستخدم البرامج بدلاً من الكلمات والصور.

من خلال عرض البيانات التي تخص كل قارئ، يمكن لتطبيق الأخبار أن يساعده على فهم الخبر بطريقة تعني له شيئًا ذا دلالة على المستوى الشخصي. كما يمكن لذلك أن يساعد القارئ في إدراك صلته الشخصية بإحدى الظواهر الوطنية الواسعة الانتشار، وفي ربط ما يعرفه بما لا يعرفه، مما يشجع على فهم عميق للمفاهيم المجردة.

نحن نميل إلى بناء تطبيقات الأخبار حين تتوفر لدينا مجموعة بيانات (أو حين نعتقد أنه من الممكن الحصول عليها) تكون ذات نطاق وطني وتتضمن أيضا مكونات دقيقة تكفي للكشف عن تفاصيل ذات معنى.

يعمل تطبيق الأخبار على سرد الخبر، لهذا فهو يحتاج إلى عنوان رئيسي وآخر فرعي وفقرة افتتاحية وموجز. وقد يصعب تمييز بعض هذه المفاهيم في البرامج التفاعلية، إلا أنه مع تدقيق النظر يمكن العثور عليها.

وينبغي أن يعمل تطبيق الأخبار على توليد المزيد من الأخبار والتقارير. وقد تم استخدام أفضل تطبيقات بروبابليكا كأحد التطبيقات الأساسية للأخبار المحلية.

على سبيل المثال، قام تطبيق الأخبار الخاص بنا 'الدولارات للأطباء' Dol-lars for Docs (http://projects.propublica.org/docdollars) للمرة الأولى

بتتبع ملايين الدولارات التي تدفعها شركات الأدوية للأطباء مقابل الاستشارات وإلقاء المحاضرات وما إلى ذلك. نجح تطبيق الأخبار الذي أنشأناه في أن يمكِّن القارئ من البحث عن طبيبه الخاص ورؤية المبالغ التي تلقاها. وقد استخدم صحافيون من مؤسسات إخبارية أخرى هذه البيانات، حيث قامت أكثر من 125 مؤسسة إخبارية محلية، بما في ذلك صحيفة بوسطن غلوب وشيكاغو تريبيون وسانت لويس بوست ديسباتش، بنشر أخبار استقصائية حول الأطباء المحليين بناء على بيانات 'الدولارات للأطباء'.

نتجت بعض تلك الأخبار المحلية عن شراكات رسمية، ولكن تم العمل على معظمها بشكل مستقل – وفي بعض الحالات، لم نعلم أن هناك من يعمل على الخبر إلا بعد نشره. وتعدُّ مثل هذه الأمور في غاية الأهمية بالنسبة لمؤسسة إخبارية صغيرة تعمل في نطاق وطني مثلنا. إذ لا يمكن أن يكون لدينا اطلاع داخلي على ما يجري في 125 مدينة، ولكن إذا كانت البيانات التي لدينا تُمكن الصحافيين الذين يتوفر لهم الاطلاع الداخلي من سرد أخبار لها تأثير، إذن فنحن نؤدي مهمتنا جيدا.

من ضمن تطبيقات الأخبار المفضلة لدي تطبيق 'خرائط لوس أنجلوس' Mapping LA (http://projects.latimes.com/mapping-la/neighbor-hoods/) في صحيفة لوس أنجلوس تايمز الذي انطلق في بادئ الأمر كخريطة تغطي العديد من الأحياء السكنية في مدينة لوس أنجلوس من خلال التكليف أو التعهيد الجماعي، ولم يكن للخريطة مجموعة من الحدود المستقلة المتفق عليها حتى وقت إطلاق التطبيق. وبعد مشروع التعهيد الجماعي الأوَلِي، استطاعت صحيفة التايمز أن تستخدم الأحياء السكنية كإطار عام في إعدادها لتقارير بيانات رائعة حول عدد من القضايا مثل معدلات الجريمة أو جودة المدارس حسب كل حي، وهي أمور لم يكن للصحيفة أن تتمكن من القيام بها من قبل. وعليه، فتطبيق 'خرائط لوس أنجلوس' ليس واسعاً ومحدداً فحسب، بل إنه أيضاً يولِّد المزيد من الأخبار التي تهم الناس وينقلها إليهم.

تتنوع الموارد اللازمة لبناء تطبيق الأخبار بشكل واسع. فلدى صحيفة نيويورك تايمز عشرات ممن يعملون على تطبيقات الأخبار والرسوم البيانية التفاعلية. ومع ذلك قامت مؤسسة 'توكينج بوينتس ميمو' Talking Points Memo

(http://polltracker.talkingpointsmemo.com/) بإنشاء تطبيق متطور لتتبع الاستطلاعات السياسية باثنين فقط من الموظفين لم يحمل أي منهما مؤهلات في علوم الكمبيوتر.

أما نحن فلا نختلف كثيراً عن معظم المبرمجين في غرف الأخبار، حيث نستخدم إحدى الطرائق المتطورة للبرمجة الرشيقة (Agile Methodology) في إنشاء التطبيقات الخاصة بنا، فنقوم بالمراجعة سريعاً ثم نعرض المسودات على الآخرين ممن يعملون معنا في غرفة الأخبار. والأهم من ذلك، أننا نعمل عن قرب مع الصحافيين ونقرأ مسوداتهم حتى النسخ الأولية منها، فيكون الأداء أقرب إلى عمل الصحافيين منه إلى المبرمجين التقليديين. وبالإضافة إلى الترميز أو كتابة الكود، نتصل بالمصادر ونجمع المعلومات ونبني الخبرات، إذ إنه من الصعب للغاية إنشاء تطبيق أخبار جيد باستخدام مواد لا نفهمها.

لماذا يجب أن تهتم غرف الأخبار بإنشاء تطبيقات أخبار قائمة على البيانات؟ لثلاثة أسباب: لأنها صحافة مميزة، ورائجة بشكل واسع حيث تعد تطبيقات الأخبار أشهر الأبواب في بروبابليكا، ولأننا إن لم نفعل، سيقوم غيرنا بإنشاء هذه التطبيقات. فكر في كل سبق صحفي قد يفوتنا! والأهم من ذلك أن على غرف الأخبار الاطمئنان لامكانية أن تفعل ذلك أيضاً. فالموضوع أسهل مما يبدو.

—— سكوت كلاين، بروبابليكا

التوضيح المرئي هو العمود الفقري لصحافة البيانات

قبل البدء في وضع رسوم بيانية أو خرائط للبيانات التي لديك، فكر لحظات في الكثير من الوظائف التي يمكن أن تؤديها كل من العناصر الرسومية الثابتة والتفاعلية في صحافتك.

في مرحلة إعداد التقارير، يمكن للتوضيح المرئي:

- المساعدة في التعرف على المواضيع والأسئلة لبقية التقارير.
- التعرف على العناصر المختلفة كالأخبار الجيدة وأحياناً الأخطاء الموجودة في البيانات.
- المساعدة في العثور على الأمثلة النموذجية.

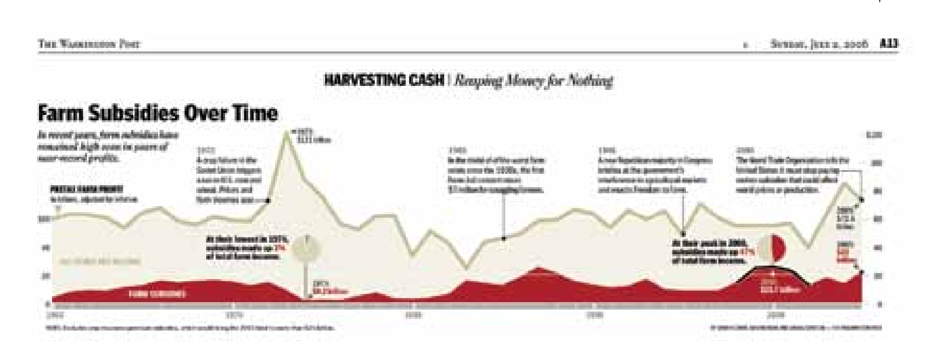
• إظهار الثغرات الموجودة في التقارير.

يؤدي التوضيح المرئي أيضاً أدواراً متعددة في النشر، مثل:

• توضيح أحد النقاط المتضمنة في الخبر بشكل أكثر إقناعاً.
• التخلص من المعلومات الفنية غير الضرورية في النص.
• إبراز الشفافية في عملية إعداد التقارير خاصة عندما يكون التوضيح المرئي تفاعلياً ويتيح الاستكشاف.

كل هذه الوظائف تفرض عليك البدء أولاً بالتوضيح المرئي في معظم تقاريرك أكنت تبدأ بالبيانات الإلكترونية أو السجلات أم لا، ولا تنظر إلى هذا الأمر على أنه خطوة منفصلة يمكن التفكير فيها بعد كتابة الجزء الأكبر من الخبر، بل ليكن هذا العمل دليلاً لك أثناء إعداد التقارير.

قد لا تتعدى البداية أحياناً وضع ملاحظاتك في شكل مرئي، وانظر على سبيل المثال الرسم البياني في الشكل (6–2) المنشور في صحيفة واشنطن بوست عام 2006.



نموذج 6–2. دعم المزارع في السنوات الماضية (واشنطن بوست)

يوضح الرسم البياني حصة دخول المزارع في مقابل الدعم الحكومي وأبرز الأحداث التي وقعت على مدى الخمسة وأربعين عاماً الماضية، وقد استغرق إعداده عدة أشهر، لأنه كان من الصعب العثور على بيانات لها تعريفات ودلالات متشابهة عبر هذه الفترة. وقد تمكنا من خلال تقصي كافة حالات الارتفاع والهبوط من المحافظة على السياق حتى انتهينا من إعداد التقرير، وكان ذلك بمثابة إنجاز أحد المهام الشاقة قبل كتابة الأخبار.

فيما يلي بعض النصائح الخاصة باستخدام التوضيح المرئي قبل البدء في استكشاف مجموعات البيانات التي لديك.

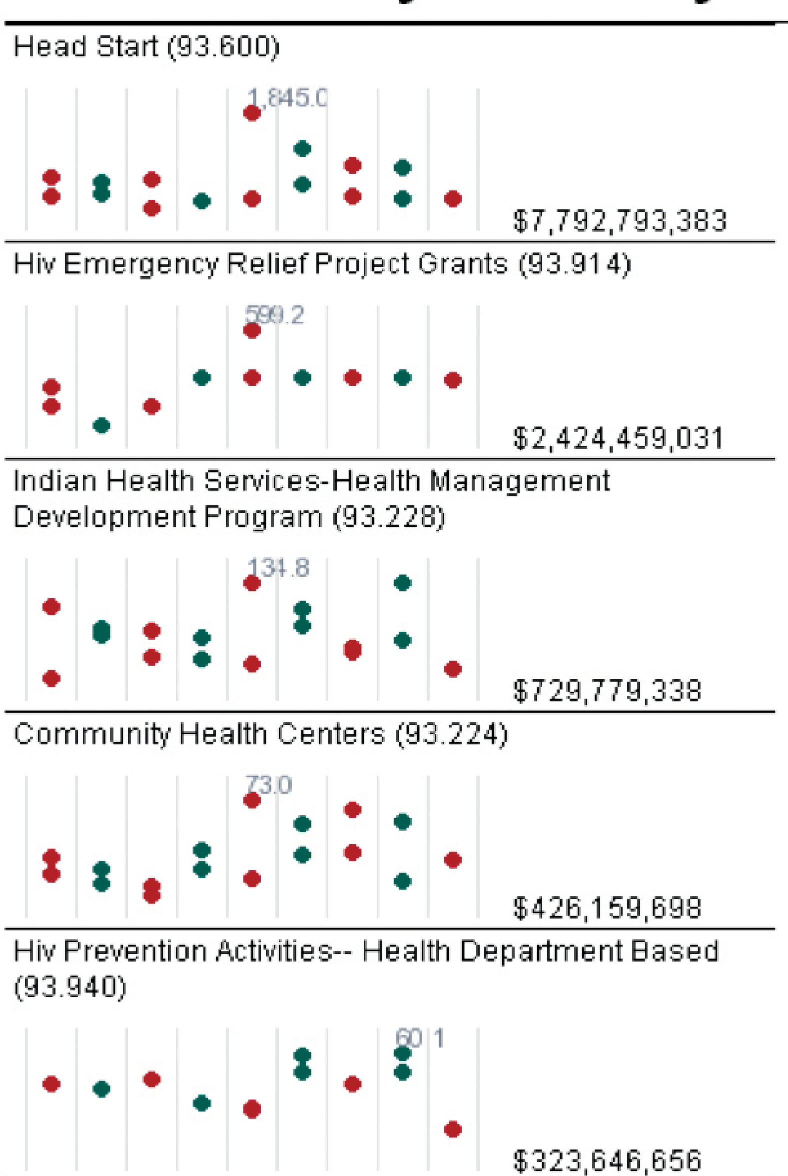
النصيحة الأولى: استخدم المُضاعِفات الصغيرة لتتأقلم بسرعة مع مجموعات البيانات الكبيرة

لقد استخدمت هذا الأسلوب في صحيفة واشنطن بوست عندما كنا نتحقق من معلومة تفيد بأن إدارة جورج بوش كانت تقدم منحاً على أساس سياسي وغير موضوعي. كانت معظم برامج هذه المعونات تتم حسب معادلة رياضية معينة، وهناك برامج أخرى يتم تمويلها منذ سنوات، ولذا ثار لدينا فضول نحو ما إذا كان بإمكاننا رؤية هذا النمط بالاطلاع على ما يقرب من 1500 نمط تقديري مختلف.

وضعت رسماً بيانياً لكل برنامج، حيث تشير النقاط الحمراء إلى عام الانتخابات الرئاسية، بينما تشير النقاط الخضراء إلى عام انتخابات الكونغرس. والمشكلة هي: وجود ارتفاع كبير بالفعل في العديد من تلك البرامج في الأشهر الستة التي سبقت الانتخابات الرئاسية – وتعبِّر عنه النقاط الحمراء وإلى جانبها قيم الارتفاع – ولكنه كان عام الانتخابات الخطأ. فبدلاً من كونه عام طلب انتخاب جورج بوش، استمر الارتفاع قبيل الانتخابات الرئاسية عام 2000 عندما كان بيل كلينتون في سدة الحكم في البيت الأبيض، ونائبه آل غور يخوض الانتخابات الرئاسية.

كان من السهل للغاية رؤية ذلك في سلسلة من الرسوم البيانية بدلاً من الأرقام والجداول، حيث أتاحت لنا هذه الصيغة التفاعلية متابعة أنواع متعددة من المنح والمناطق والهيئات ذات الصلة. وتعد الخرائط التي تستخدم فيها مضاعفات صغيرة أحد الوسائل الفعالة لعرض القيم الزمانية والمكانية على صورة ثابتة تسهل مقارنتها، وتكون أحياناً أسهل حتى من الأشكال التفاعلية.

استُخدم برنامج قصير مكتوب بلغة البرمجة PHP في تطوير هذا النموذج، والآن أصبح إنجازه أسهل بكثير عما سبق من خلال استخدام الرسوم البيانية المصغرة في برنامج إكسل 2007 و2010. وقد ابتكر خبير التوضيح المرئي إدوارد توفتي تلك 'الرسوم البيانية المكثفة والبسيطة والمشابهة للكلمات' لتوصيل



نموذج 6–3. منح وزارة الصحة والخدمات الإنسانية: الخطوط البيانية تساعد في تسليط الضوء على الخبر (واشنطن بوست)

المعلومات من خلال إلقاء نظرة سريعة على مجموعة بيانات هائلة الحجم. والآن تجد هذه النماذج في كل مكان، بدءاً من الرسوم البيانية الصغيرة لأسهم الأسواق المالية وصولاً إلى سجلات الفوز والخسارة الرياضية.

النصيحة الثانية: انظر إلى البيانات التي لديك من جميع الاتجاهات

ليس ثمة طريقة خطأ للنظر في خبر ما أو مجموعة من البيانات عند محاولة فهم القصة الإخبارية، لذلك حاول تناولها بشتى الطرق التي تخطر على بالك، وستخرج بوجهة نظر مختلفة. فإذا كنت بصدد إعداد تقرير عن الجرائم، فقد تطالع مجموعة من الرسوم البيانية التي توضح التغيرات في جرائم العنف خلال عام واحد، أو في مجموعة أخرى توضح التغير في معدلاتها بنسبة مئوية، أو تقارن بوقوعها في مدن أخرى، أو توضح التغيرات في معدلاتها عبر فترة من الزمن، مع ضرورة استخدام الأرقام والنسب والمؤشرات المجردة.

انظر إليها بمقاييس مختلفة، وحاول أن تتبع قاعدة أن محور (X) يجب أن يكون صفراً. ثم اكسر تلك القاعدة لترى إن كان بإمكانك معرفة المزيد. جرب اللوغاريتمات والجذور التربيعية مع البيانات ذات التقسيمات الغريبة.

احرص على الاستفادة من الأبحاث التي أجريت حول الإدراك البصري. فقد أظهرت تجارب ويليام كليفلاند أن العين ترى التغيير في الصورة عندما تنظر إلى المنحنى البياني مائلا بزاوية قدرها خمس وأربعون درجة في المتوسط. وهذا يعني أهمية تجاهل النصائح التي تحض دائماً على البدء من الصفر، مع ضرورة العمل نحو الرسوم البيانية التي يسهل استيعابها بشكل كبير. هناك أبحاث أخرى في علم الأوبئة توصي بإيجاد مستوى منشود وجعله حدّا للرسم البياني الخاص بك. كل واحدة من هذه الطرق تعين على رؤية البيانات بطريقة مختلفة، وعندما تتوقف تلك الطرق عن إخبارك بشيء جديد، تدرك عندئذ أنك أنهيت العمل.

النصيحة الثالثة: لا تفترض شيئاً

والآن بعد أن نظرت في البيانات التي لديك بشتّى الطرق، ربما تعثر على سجلات لا تبدو صحيحة أو قد لا تفهم ما تعنيه لأول وهلة، أو أن هناك بعض

القيم المختلفة التي تبدو كأنها أخطاء إملائية، أو أن هناك بعض الاتجاهات التي تبدو معكوسة.

وإذا أردت أن تنشر أي شيء على هيئة توضيح مرئي منشور بناء على استكشافاتك الأولية، يتوجب عليك حل هذه الإشكالات والامتناع عن تقديم الافتراضات، لأنها إما أن تكون أخباراً مثيرة للاهتمام أو أخطاءً، أي أنها إما تحديات للمفاهيم السائدة أو سوء فهم.

لا يُستغرب أن تقدم الحكومات المحلية جداول بيانية مليئة بالأخطاء، ووارد جداً أن يساء فهم أحد التعبيرات الخاصة التي تستخدمها الحكومة في مجموعة البيانات.

أولاً، عاود الرجوع إلى العمل الذي قمت به، وأسأل نفسك: هل قرأت الوثائق والتحذيرات؟ وهل المشكلة موجودة في النسخة الأصلية من البيانات؟ فإذا كان كل شيء من ناحيتك سليما، فقد حان الوقت لإجراء بعض الاتصالات، وسيتوجب عليك إيجاد حل لتلك المشكلة إن كنت تريد استخدام هذه البيانات، فمن الأفضل أن تبدأ على الفور.

إضافة إلى ذلك، ليس كل خطأ مهماً. ففي سجلات تمويل الحملات الانتخابية، من الشائع أن تجد في قاعدة بيانات تتألف من 100,000 سجل المئات من الرموز البريدية التي ليس لها وجود. وطالما أنها لا توجد كلها في نفس المدينة أو لمرشح واحد، فليس ثمة مشكلة في سجل البيانات الذي قد يكون غير دقيق أحياناً.

السؤال الذي تطرحه على نفسك هو: هل سيحصل القراء على رؤية دقيقة لما تتضمنه البيانات إذا قررت استخدام تلك البيانات؟

النصيحة الرابعة: تجنب الإفراط في الدقة

على الجانب المقابل لعدم طرح الأسئلة الكافية هناك جانب الإفراط في العناية بالدقة قبل أن يحين أوان الاهتمام بها لهذه الدرجة. فمن الضروري أن تكون رسوماتك البيانية الاستكشافية صحيحة بشكل عام، ولكن لا تقلق إن ظلت لديك عدة مستويات من التقريب أو إذا لم تصل نسبة الدقة إلى 100٪ أو إن كنت

تفتقد بيانات سنة أو سنتين من أصل 20 سنة، لأن هذا جزء من عملية الاستكشاف، وستظل قادرا على رؤية الاتجاهات الرئيسية لتدرك الأشياء التي ينبغي لك جمعها قبل أن يحين وقت النشر.

في الواقع، قد تضطر إلى إزالة الأسماء وعلامات المقاييس، كما هو الحال في الرسوم البيانية المذكورة أعلاه، كي تتمكن من استيعاب البيانات بشكل عام.

النصيحة الخامسة: رتب القضايا والأحداث ترتيباً زمنياً

ابدأ بترتيب القضايا والأحداث الرئيسية ترتيباً زمنياً في مستهل أي خبر معقد. ويمكنك الاستعانة بأحد مستندات إكسل أو وورد أو أي أداة خاصة مثل TimeFlow للقيام بهذه المهمة، وستجد في مرحلة معينة مجموعة بيانات يمكنك أن تضعها في الخلفية، وستظهر لك من خلال قراءتها من حين لآخر بعض الفجوات التي ينبغي سدها في التقرير.

النصيحة السادسة: اجتمع مع قسم الغرافيكس مبكراً وباستمرار

احرص على تبادل الأفكار مع الفنانين والمصممين في غرفة الأخبار حول الرسوم البيانية الممكنة. ستكون لديهم طرق مبتكرة للنظر في البيانات التي لديك، وكذلك مقترحات خاصة بطريقة عملها بشكل تفاعلي، بالإضافة إلى معرفة كيفية ربط البيانات بالأخبار. سيسهل عليك كثيرا عند كتابة الخبر أن تكون لديك معرفة مبكرة بالبيانات التي ينبغي جمعها؛ أو أن تقوم مبكرا بتنبيه فريق العمل بتعذر إنشاء أحد الرسومات البيانية بسبب عدم إمكانية جمع هذه البيانات.

نصائح للنشر

قد تستغرق مرحلة الاستكشاف لديك بضعة أيام أو ساعات فقط، وقد يكون الخبر قد استغرق شهوراً لإعداده. وعندما يحين وقت النشر، ينبغي التركيز على جانبين مهمين:

هل تتذكر تلك السنة المفقودة في استكشافك الأوّلي للبيانات؟ فجأة ستدرك انه لا يمكنك أن تتقدم قيد أنملة بدونها. وكل تلك البيانات الرديئة التي تجاهلتها أثناء إعداد تقريرك؟ ستعود لتطاردك من جديد.

والسبب أنه ليس من الممكن الكتابة باستخدام بيانات رديئة؛ وأما في الرسوم البيانية، فإما أن يكون لديك كل ما تحتاجه أو لا، وليس هناك حل وسط.

ملاءمة الجهد المبذول في جمع البيانات للرسم البياني التفاعلي

ليس ثمة ما يمكن أن تخفيه في الرسم البياني التفاعلي. فإذا كنت تنوي فعلاً تمكين القراء من استكشاف البيانات بالطريقة التي يرغبونها، لا بد أن يُعبّر كل عنصر في البيانات بصدق عن حقيقته وان يقدم ما هو متوقع منه، لأن المستخدمين يمكنهم العثور على الأخطاء في أي وقت، وقد يطاردك شبح هذا الأمر لأشهر أو سنوات.

فإذا كنت بصدد بناء قاعدة البيانات الخاصة بك، فيُتوقع منك إجراء التدقيق اللغوي والتأكد من الحقائق وتحرير قاعدة البيانات بأكملها. وإذا كنت تستخدم السجلات الحكومية، فعليك أن تحدد القدر اللازم من الاختبارات والفحوص التي ستقوم بها، وما الذي ستفعله عند العثور على خطأ لا يمكن تفاديه.

تصميم الرسومات لتناسب نوعين من القراء

يشترط في الرسم البياني أن يرضي نوعين مختلفين من القراء، سواء أكان هذا الرسم ذا خاصية تفاعلية مستقلة أم توضيحاً مرئياً ثابتاً يعمل على طول الخبر، وأن يسهل استيعابه بمجرد النظر إليه مع وجود قدر من التعقيد ليقدم شيئاً مثيراً للاهتمام لمن يريد التحقق أكثر. وإذا اخترت أن يكون الرسم البياني تفاعلياً، تأكد من حصول القراء على أكثر من مجرد أرقام أو أسماء مفردة.

أعرض فكرة واحدة ثم قم بتبسيطها

احرص على إبراز شيء واحد ترغب في إطلاع الناس عليه، وحدد الفكرة الأهم التي تود إيصالها للقارئ مع استبعاد كافة الأمور الأخرى. وقد يعني هذا في كثير من الحالات حذف بعض المعلومات حتى في الوقت الذي يسمح لك فيه الإنترنت بتوفير كافة المعلومات. وما لم يكن الغرض الأساسي إبراز شفافية التقارير الإخبارية المنشورة، فإن معظم التفاصيل التي جمعتها في الجدول أو الإطار الزمني ليست بالغة الأهمية، بل ستبدو مخيفة في الرسومات البيانية الثابتة ومملة في الرسومات البيانية التفاعلية.

استخدام التوضيح المرئي لسرد الأخبار

يستحق التوضيح المرئي للبيانات النظر فيه لعدة أسباب، منها روعة جماله وقدرته الفائقة على جذب الانتباه، وهذا بمثابة عملة اجتماعية قيّمة تعزز مشاركة القراء وجذبهم. ليس هذا فحسب، بل توظف إحدى الأدوات المعرفية القوية، وهي أن نصف المخ البشري مخصص كلياً لمعالجة المعلومات المرئية. فعندما تقدم للمستخدم المعلومات في شكل رسم بياني، فأنت تصل إليه عن طريق أكبر حيز في العقل. ويستطيع التوضيح المرئي للبيانات المُصمم بشكل جيد أن يعطي المشاهدين انطباعاً فورياً وعميقاً وأن يختصر الأجزاء المتناثرة للخبر المعقد ليصل مباشرة إلى الخلاصة.

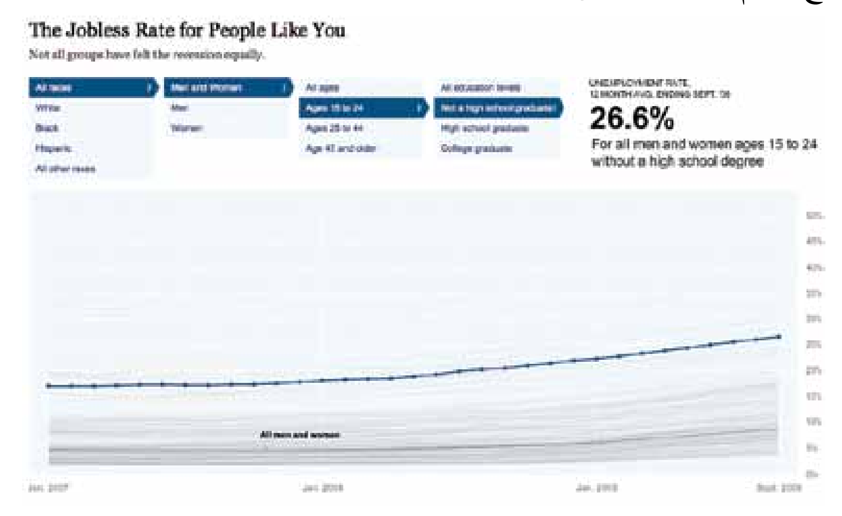
وبخلاف الوسائل المرئية الأخرى – مثل التصوير الفوتوغرافي وتصوير الفيديو – فالتوضيح المرئي للبيانات متأصل في الحقائق القابلة للإدراك. وهو جذاب من الناحية الجمالية وفي نفس الوقت أقل عاطفية وأكثر اهتماماً بتسليط الضوء على الموضوع بدلاً من الإثارة. وفي عصر الإعلام الضيق الأفق الذي عادة ما يصاغ لتوجيه الجماهير نحو وجهة نظر معينة، يقدم التوضيح المرئي للبيانات – وصحافة البيانات عموماً – فرصة مثيرة لسرد الأخبار المبنية على الحقائق فوق كل شيء وليس على التعصب.

بالإضافة إلى ذلك، يمكن للتوضيح المرئي للبيانات، كغيره من أشكال أخرى من الصحافة السردية، أن يكون أداة فعالة لنقل كل من الأخبار العاجلة من خلال نقل المعلومات الجديدة سريعاً مثل موقع الحادث وعدد الإصابات؛ وكذلك نقل التقارير غير السياسية من خلال التعمق في الموضوع وتقديم تصور مبتكر يساعدك على رؤية شيء مألوف من منظور جديد تماماً.

رؤية المألوف بشكل جديد

في الواقع، تظهر قدرة التوضيح المرئي للبيانات على اختبار الأفكار السائدة في الرسم البياني التفاعلي (http://nyti.ms/employment-lines) الذي نشرته صحيفة نيويورك تايمز في أواخر عام 2009، أي بعد عام من بدء الأزمة الاقتصادية العالمية. فمع ارتفاع معدلات البطالة في الولايات المتحدة إلى حوالي 9٪، تمكن

المستخدمون من فرز سكان الولايات المتحدة بواسطة أدوات تصنيف ديموغرافية وتعليمية عديدة ليروا التفاوت الكبير في المعدلات. وقد اتضح أن المعدلات تراوحت بين أقل من 4٪ للنساء في منتصف العمر والحاصلات على درجات علمية متقدمة وبين حوالي نصف الشباب السود الذين لم يكملوا المرحلة الثانوية. ولم يكن هذا التفاوت جديداً، بل يوضح حقيقة أكدت عليها الخطوط البيانية التي توضح القيم التاريخية لكل من تلك المجموعات.



نموذج 6–4. معدل البطالة لأناس مثلك (نيويورك تايمز)

وحتى بعد أن تتوقف عن النظر إليه، ينفذ التوضيح المرئي للبيانات إلى عقلك ليترك في الذهن صورة دائمة لإحدى الحقائق أو التوجهات أو العمليات. كم شخصا رأى الرسم المتحرك الذي نشره باحثو تسونامي (http://1.usa.gov/tsunami-animation) في ديسمبر 2004، والذي أظهر الأمواج المتتالية التي نتجت عن زلزال إندونيسيا وجاءت من المحيط الهندي مهددة ملايين السكان الساحليين في جنوب آسيا وشرق أفريقيا؟

تستطيع التوضيحات المرئية للبيانات وما يصاحبها من اللمسات الجمالية أن تصبح معايير ثقافية أيضاً، كتمثيل الانقسامات السياسية العميقة في الولايات المتحدة بعد انتخابات عام 2000 وعام 2004، عندما غطت الولايات 'الحمراء' التي صوتت للحزب الجمهوري وسط البلاد، بينما تجمعت الولايات 'الزرقاء'

التي صوتت للحزب الديموقراطي في الشمال الشرقي وأقصى الغرب. وليس ثمة مشكلة في قيام شبكات البث الرئيسية في الإعلام الأمريكي قبل عام 2000 باستخدام اللون الأحمر والأزرق والتبديل بينهما لتمثيل كل حزب، حتى إن بعضها اختار أن يبدل كل أربعة أعوام. وهذا يفسر ذكريات بعض الأمريكيين لنصر رونالد ريغان الساحق بتسع وأربعين ولاية 'زرقاء' (http://bit.ly/infobarrel-map) صوتت للجمهوريين عام 1984.

ومقابل كل رسم بياني ينتج عنه أحد الكليشيهات المرئية، يأتي آخر ليقدم شهادة واقعية قوية، مثل الخريطة التي نشرتها صحيفة نيويورك تايمز عام 2006 (http://nyti.ms/diaspora-graphic) باستخدام دوائر مختلفة الأحجام تعرض أين يعيش الآن مئات الآلاف ممن تم إخلاؤهم من نيو أورليانز والمشتتون في أنحاء القارة من خلال مزيج من العلاقات الشخصية وبرامج إعادة التوطين. هل سيتمكن هؤلاء 'العالقون' الذين تم إخلاؤهم من العودة إلى منازلهم يوماً ما؟

والآن بعد أن ألقينا الضوء على قوة التوضيح المرئي للبيانات، فمن الإنصاف أن نسأل: متى ينبغي أن نستخدمه؟ ومتى ينبغي ألا نستخدمه؟

متى يستخدم التوضيح المرئي للبيانات؟

سننظر أولاً في بعض الأمثلة على الحالات التي قد يكون من المفيد فيها استخدام التوضيح المرئي للبيانات ليساعدك على سرد الخبر للقراء.

عرض التغيرات عبر الزمن

لعل من أكثر الاستخدامات الشائعة للتوضيح المرئي للبيانات - كما تظهر في الخطوط البيانية البسيطة - عرض التغيرات في القيم عبر الزمن. ومن الأمثلة الجيدة على ذلك نمو عدد سكان الصين منذ عام 1960 (http://bit.ly/google-vhi-na-population) أو الارتفاع الكبير في معدلات البطالة منذ الانهيار الاقتصادي عام 2008. ويكون التوضيح المرئي للبيانات قوياً أيضاً في عرض التغيرات عبر الزمن من خلال أشكال بيانية أخرى. فقد استخدم الباحث البرتغالي بيدرو م. كروث رسوماً بيانية بدوائر متحركة ليعرض بشكل كبير انهيار إمبراطوريات أوروبا

الغربية (http://pmcruz.com/visual-experiments/visualizing-em-pires) منذ بدايات القرن التاسع عشر. ومع ترتيب حجم الدوائر حسب تعداد السكان، تبرز بريطانيا وفرنسا وإسبانيا كالفقاعات كلما حصلت أراضيها الخارجية على الاستقلال. وهكذا انهارت المكسيك والبرازيل وأستراليا والهند، وغيرها في الانتظار، وانهارت العديد من المستعمرات الإفريقية في بدايات الستينيات مما كاد يسبب تصدعات في قلب فرنسا.

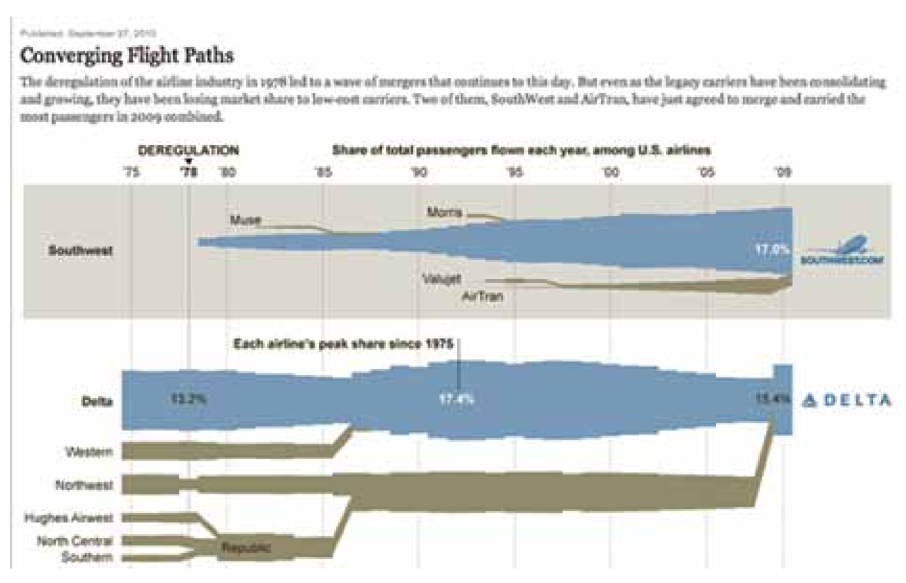
يعرض الرسم البياني الذي نشرته وول ستريت جورنال (http://on.wsj.com/tech-empire) عدد الأشهر التي قضاها 100 من رواد الأعمال لتحقيق رقم معين من العائدات بلغ 50 مليون دولار. أُجريت المقارنة باستخدام أداة الرسم البياني وتحليل البيانات المجانية Tableau Public، وظهرت المقارنة في شكل الخطوط النفاثة التي تتشكل خلف مجموعة من الطائرات عند الإقلاع، منها السريع والبطيء والثخين، رسمت فوق بعضها البعض.

وبمناسبة الحديث عن الطائرات، هناك رسم بياني آخر مثير للاهتمام يعرض التغيرات عبر الزمن ويوضح حصة كبرى شركات الطيران الأمريكية في السوق (http://nyti.ms/airline-merger) خلال عدة عقود من توحيد الصناعة. وبعد أن رفعت إدارة كارتر القيود عن طيران نقل الركاب، أدى عدد كبير من عمليات الاستحواذ والدمج الممولة بالديون إلى اندماج شركات الطيران الإقليمية الصغيرة إلى شركات طيران وطنية كما يوضح هذا الرسم البياني الذي نشرته صحيفة نيويورك تايمز.

بما أن معظم القراء العاديين يرون أن محور x الأفقي في المخطط البياني يمثل الزمن، فأحياناً يكون من السهل الاعتقاد بأن كافة التوضيحات المرئية ينبغي أن تعرض تغيرات عبر الزمن.

مقارنة القِيَم

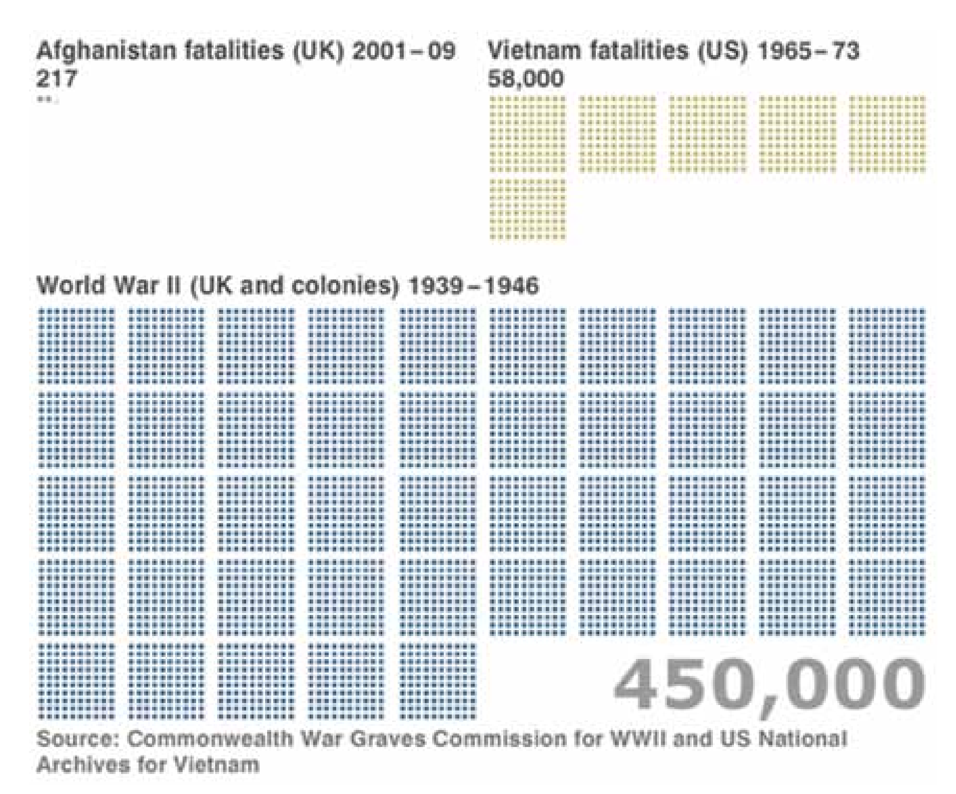
يتألق التوضيح المرئي للبيانات أيضاً فيما يتعلق بمساعدة القراء على مقارنة قيمتين منفصلتين أو أكثر، سواء كان ذلك لرؤية سياق الخسارة المأساوية للنساء والرجال من الجنود في صراعات العراق وأفغانستان (من خلال مقارنتهم بالآلاف



نموذج 6–5. تغطية مسارات الطيران (نيويورك تايمز)

الذين قُتلوا في فيتنام والملايين الذين ماتوا في الحرب العالمية الثانية، كما فعلت شبكة بي. بي. سي. في عرض الصور المتحركة (http://bbc.in/animated-slideshow) المصاحبة لقاعدة بيانات المصابين الخاصة بهم)؛ أو عندما عرضت ناشيونال جيوغرافيك باستخدام رسم بياني في غاية البساطة (http://bit.ly/ngm-hearts) لنسبة احتمال الوفاة بأمراض القلب (1 من 5) أو بالسكتة الدماغية (1 من 24) مقابل احتمال الوفاة في حوادث الطائرات مثلاً (1 من 5,051) أو من لسع النحل (1 من 56,789) وذلك من خلال عرض الاحتمالات النسبية للوفاة (يهيمن عليها جميعاً قوس كبير يمثل احتمالات الوفاة بشكل عام: 1 من 1!).

طورت شبكة بي. بي. سي. أيضاً، بالتعاون مع وكالة بيرغ ديزاين، الموقع الإلكتروني 'أبعاد' Dimensions (http://bit.ly/ngm-hearts)، الذي يتيح لك وضع معالم أهم الأحداث العالمية على خريطة جوجل في المجتمع الذي تعيش فيه، مثل التسرب النفطي في خليج المكسيك أو فيضانات باكستان (http://howbigreally.com/).

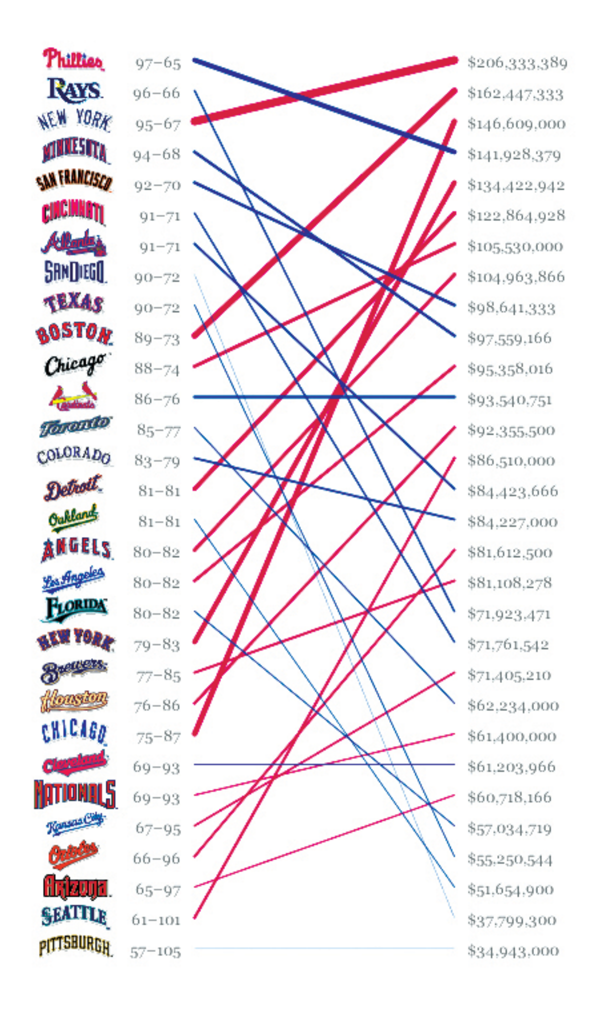


نموذج 6–6. حساب التكلفة البشرية للحرب (بي. بي. سي.)

عرض الصلات والروابط

لم يجعل تشغيل فرنسا للقطارات فائقة السرعة عام 1981 البلد أصغر مساحة بالمعنى الحرفي، ولكن هناك تمثيلاً مرئياً ذكياً يوضح أن الزمن الذي يستغرقه القطار الجديد الآن للوصول إلى وجهات مختلفة أقل مما كان يستغرقه القطار العادي. تم وضع شبكة فوق خريطة فرنسا وظهرت بشكل مربعات سليمة في صورة "ما قبل القطار السريع"، بينما ظهرت بشكل مضغوط مركزياً حول باريس في صورة "ما بعد القطار السريع". وقد أظهر هذا أن الرحلات الخارجة من باريس ليست هي الأقرب فحسب، بل أن أعلى مقدار من توفير الوقت يحدث في الجزء الأول من الرحلة قبل أن يصل القطار إلى السكك الحديدية غير المطورة ويضطر إلى السير ببطء.

للمقارنة بين متغيرين منفصلين، انظر إلى رسم ’بين فراي‘ البياني الذي يقيّم أداء فرق البيسبول في الدوري بالنسبة إلى رواتبهم. في العمود الأيسر، تم تصنيف الفرق وفقاً لإنجازاتها السابقة حتى اليوم، بينما في العمود الأيمن نرى مجموع رواتب اللاعبين. ويربط بين القيمتين خطٌ أحمر (أداء ضعيف) وخطٌ أزرق (أداء قوي)، مما يجعلنا ندرك من مُلاك الفرق يندم على ضعف أداء لاعبيهم ممن يتقاضون أجوراً عالية. إضافة إلى ذلك، يمنحنا تعقب التتابع الزمني للفرق متابعة حية لمنافسات الفوز في الموسم الحالي حتى النهاية.



نموذج 6–7. الرواتب مقابل الأداء (بين فراي)

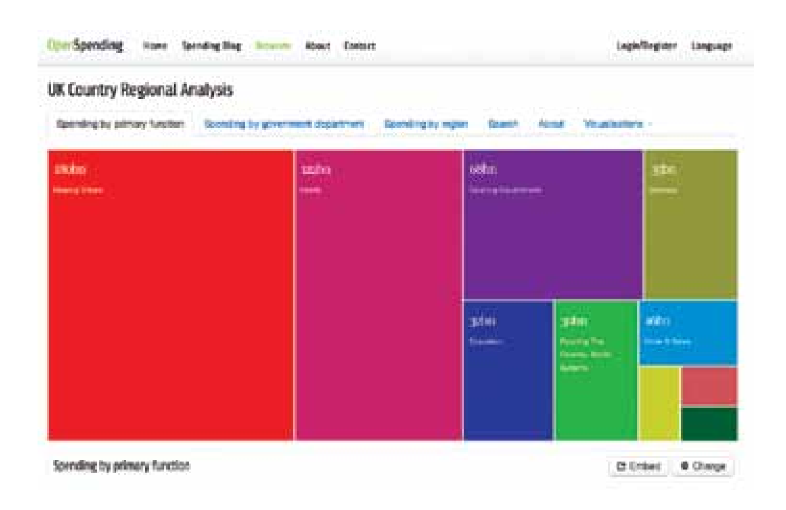
تقوم مخططات التدفق أيضاً بتشفير المعلومات في خطوط متصلة بشكل مشابه للرسم البياني لنقاط الاتصال، عادة بالسماكة أو بالألوان. فمثلاً، مع أزمة منطقة اليورو وعدم قدرة العديد من الدول الأعضاء على سداد ديونها، سعت صحيفة نيويورك تايمز إلى حل شبكة الاقتراض (http://nyti.ms/eurozone–crisis) التي تربط الدول الأعضاء في الاتحاد الأوروبي بشركائهم التجاريين عبر المحيط الأطلسي وفي آسيا. في إحدى 'حالات' التوضيح المرئي، يعكس عرض الخط حجم القروض التي تنتقل من دولة إلى أخرى، بينما حيث توجد زيادة في اللون الأصفر أو البرتقالي فإن هذا يشير إلى مدى كونه 'مقلقا'، أي أنه من غير المرجح أن يُسدد القرض.

على صعيد أكثر إشراقاً، نشرت ناشونال جيوغرافيك رسماً بيانياً مخادعاً في بساطته يعرض الصلات بين ثلاث مدن أمريكية: نيويورك وشيكاغو ولوس أنجلوس (http://bit.ly/sankey–wine) وبين أهم المناطق المنتجة للنبيذ، ويوضح كيف أن أساليب نقل المنتج من كل من المصادر قد تؤدي إلى بصمة كربونية (أي درجة تلوث للهواء) تختلف بشكل جذري عن الأخرى، مما يعني مثلاً أن اختيار سكان نيويورك لنبيذ وارد من بوردو في فرنسا قد يكون أكثر محافظة على البيئة من اختيارهم لنبيذ قادم من ولاية كاليفورنيا.

إن "Sourcemap" – وهو مشروع بدأ في كلية إدارة الأعمال في معهد ماساتشوستس للتكنولوجيا MIT – يستخدم مخططات التدفق ليلقي نظرة دقيقة جداً على المشتريات العالمية من المنتجات المصنعة ومحتوياتها وموادها الخام. وبفضل الكثير من الأبحاث المكثفة، يستطيع المستخدم الآن أن يبحث عن المنتجات التي تتراوح بين أحذية من ماركة Ecco (http://sourcemap.com/view/1760) إلى عصير البرتقال (http://sourcemap.com/view/1011) وأن يعرف مصدرها من أي مكان في العالم وما هي بصمتها الكربونية.

عرض التسلسل الهرمي

في عام 1991، ابتكر الباحث 'بين شنايدرمان' شكلاً جديداً من أشكال التوضيح المرئي يُدعى 'Treemap' (http://www.cs.umd.edu/hcil/treemap-history/) ويتكون من عدة مربعات مستقرة داخل بعضها البعض. تشير مساحة أي مربع إلى الكمية التي يمثلها، سواء بذاته أو كمجموع لمحتوياته. وسواء أكنت تضع توضيحاً مرئياً لميزانية وطنية حسب كل وكالة رئيسية وفرعية (http://openspending.org/)، أو لبورصة سوق الأسهم حسب كل قطاع وكل شركة، أو للغة برمجة حسب التصنيفات الرئيسية والفرعية، فإن Treemap هو وصلة إنترفيس محكمة وسهلة الاستخدام لوضع خريطة كيان كامل بمحتوياته. هناك شكل فعّال آخر يُدعى 'dendrogram' ويبدو كأي هيكل تنظيمي عادي حيث تستمر الفئات الفرعية في التشعب من جذع أصلي واحد.



نموذج 6–8. OpenSpending.org (مؤسسة المعرفة المفتوحة)

تصفح قواعد البيانات الكبيرة

أحياناً يكون التوضيح المرئي للبيانات فعّالاً جداً في عرض المعلومات المألوفة بشكل جديد تماماً، ولكن ماذا يحدث عندما تكون لديك معلومات جديدة كليا ويريد الناس تصفّحها؟ إن عصر البيانات يجلب معه اكتشافات جديدة ومذهلة

يوميا تقريباً، بدءاً من التحليل الجغرافي المذهل الذي ابتكره إيريك فيشر للقطات الفوتوغرافية على موقع فليكر (http://bit.ly/flicker-analysis)، وصولاً إلى نشر نيوريوك لآلاف من تقييمات المعلمين التي كانت سرية من قبل (http://projects.wsj.com/nyc-teachers/).

تتجلى قوة مجموعات البيانات تلك عندما يتمكن المستخدمون من البحث والتنقيب في المعلومات التي تهمهم.

في بدايات عام 2010، حصلت صحيفة نيويورك تايمز على سجلات موقع نيتفليكس Netflix الخاصة عادة والتي توضح في أي المناطق تؤجر أي الأفلام في أغلب الأحيان. وعلى الرغم من أن موقع نيتفليكس رفض الإعلان عن الأرقام الخام، فقد نشرت نيويورك تايمز قاعدة بيانات تفاعلية جيدة (http://nyti.ms/interactive-database) تتيح للمستخدمين تصفح أكثر من مئة فيلم يتم استئجارها في 12 مدينة كبرى في الولايات المتحدة، مصنفة حسب الرمز البريدي. وقد أتاحت الخريطة الملونة للمستخدمين أن يتصفحوها سريعاً ويروا مكان شعبية أي فيلم معين.

وبنهاية نفس العام، نشرت صحيفة التايمز نتائج تعداد السكان الذي يتم كل عشر سنوات (http://nyti.ms/census-explorer) بعد ساعات قليلة من الإعلان عنها. وقد قدمت وصلة إنترفيس المستخدم، التي تم بناؤها باستخدام Adobe Flash، عددا من خيارات التوضيح المرئي وسمحت للمستخدمين بتصفح كل مجموعة تعداد سكاني في البلاد (من ضمن 8.2 مليون مجموعة) ليروا توزيع السكان حسب العرق والدخل والتعليم. وقد كانت البيانات دقيقة لدرجة أنك إذا كنت تتصفح مجموعة البيانات في الساعات الأولى بعد نشرها، فربما ستتساءل ما إن كنت أول شخص في العالم يطلع على هذا الركن من مجموعة البيانات.

وهناك استخدامات مشابهة وجديرة بالثناء للتوضيح المرئي كواجهة أمامية لقاعدة البيانات، من بينها تحقيق بي. بي. سي. في الوفيات الناتجة عن الحوادث المرورية (http://bbc.in/road-deaths)، ومنها أيضاً محاولات عديدة لترتيب دفعات البيانات الكبيرة بشكل سريع مثل نشر ويكيليكس لسجلات حرب العراق وأفغانستان.

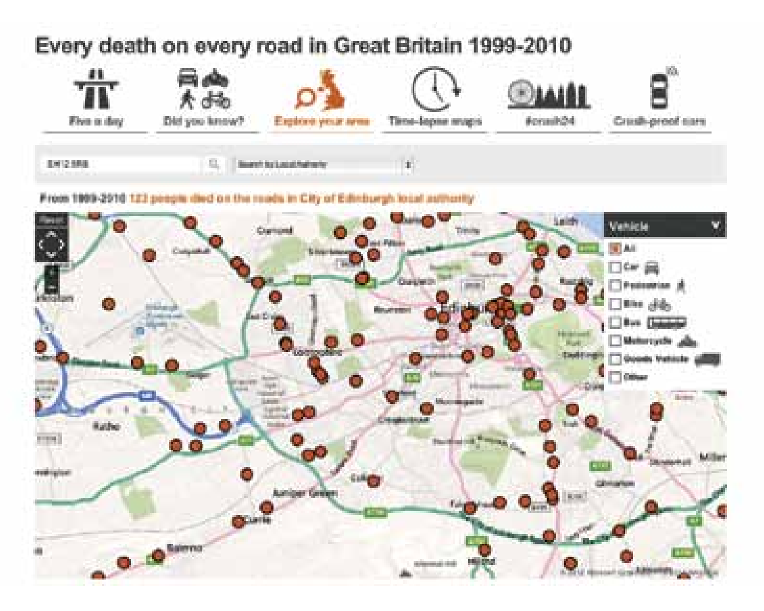
قاعدة الخمسة وستين ألفاً

عند تلقّي أول دفعة من بيانات سجلات حرب أفغانستان من ويكيليكس، بدأ الفريق المسؤول عن معالجتها بالتحدث عن مدى حماسهم للحصول على 65 ألف سجل حربي.

وقد تنبه لذلك فوراً هؤلاء الذين لديهم خبرة في التعامل مع برنامج إكسل. فبفضل قيود سابقة في التعامل مع الصفوف، لا تعالج أداة الاستيراد في إكسل أكثر من 65,536 سجل، فكأن المتحمسين إنما جاءوا برقمهم كي يناسب قدرة إكسل. وفي تلك الحالة، تبين أن هناك 25,000 سجل مفقودين فقط!

الهدف من تلك القصة (إضافة إلى تجنب استخدام إكسل في مثل تلك الحالات) هو ضرورة أن ترتاب دائماً من أي شخص يتفاخر بـوجود 65,000 صف من البيانات، فكأنه أورد الرقم من معرفته بقدرة إكسل لا من حقيقة ما لديه من بيانات.

—— ألاستير دانت، صحيفة الغارديان



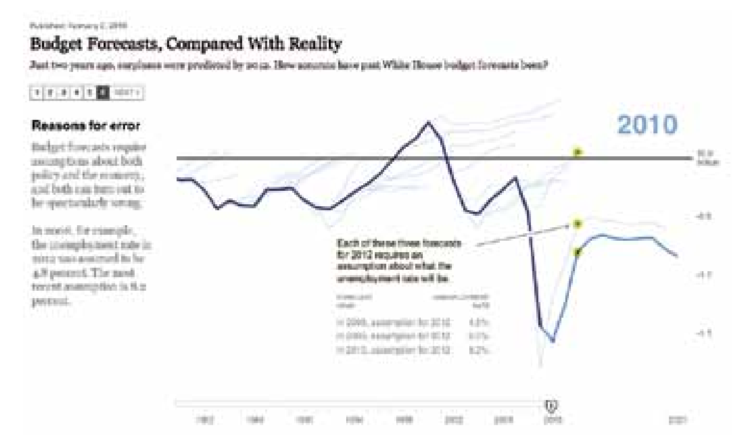
نموذج 6–9. كل حادثة وفاة على كل طريق في بريطانيا من 1999 إلى 2010 (بي. بي. سي.)

تصوّر نتائج بديلة

في صحيفة نيويورك تايمز، توضح أماندا كوكس في الرسم البياني 'procupine chart' التوقعات المتفائلة بشكل مبالغ فيه للعجز في الولايات المتحدة عبر السنوات، وكيف يكون ما حدث في بعض الأحيان أقل إثارة للاهتمام مما لم يحدث. كما توضح خطوطها البيانية التي تظهر الارتفاع الكبير في عجز الموازنة بعد استمرار الحروب أكثر من عشر سنوات وما قدم خلالها من إعفاءات ضريبية، كيف يمكن للتوقعات المستقبلية أن تكون غير واقعية على الإطلاق.

قام بريت فيكتور، وهو مصمم وصلات إنترفيس قديم لدى شركة آبل ومؤسس نظرية 'التخلص من الرياضيات' 'Kill Math' للتوضيح المرئي ونقل المعلومات الكمية، بوضع نموذج أولي لنوع من الوثائق التفاعلية (http://worrydream.com/#!/TenBrighterIdeas). في المثال الذي استخدمه، تتضمن أفكار حفظ الطاقة مواضع متسعة قابلة للتنقيح والتحسين والتطوير، حيث يمكن لخطوة بسيطة مثل إطفاء الأنوار في الغرف الخالية أن توفر للأمريكيين الطاقة التي تنتجها من 2 إلى 40 محطة توليد تُدار بالفحم. ويؤدي تغيير النسبة في منتصف الفقرة إلى تحديث النص في باقي الصفحة وفقاً لذلك!

للمزيد من الأمثلة والاقتراحات، إليك قائمة بالاستخدامات المختلفة للتوضيح المرئي والخرائط والرسوم البيانية التفاعلية جمعها ماثيو إيريكسون الذي يعمل في صحيفة نيويورك تايمز (http://bit.ly/ericson–links).



نموذج 6–10. تنبؤات الموازنة مقارنة بالواقع (نيويورك تايمز)

متى ينبغي الامتناع عن استخدام التوضيح المرئي للبيانات

في نهاية المطاف، يعتمد التوضيح المرئي الفعّال على معلومات جيدة ونظيفة ودقيقة وذات معنى. وكما أن الاقتباسات الجيدة، والحقائق والأوصاف تدعم الصحافة التقليدية الجيدة، فجودة التوضيح المرئي تعتمد فقط على البيانات التي تدعمه.

عندما يمكن سرد الخبر بشكل أفضل من خلال النص أو الوسائط المتعددة

أحياناً لا تمكننا البيانات وحدها من سرد الخبر بالطريقة المثلى. وعلى الرغم من أن الرسم البياني البسيط الذي يوضح اتجاهاً أو إحصائية يمكن أن يكون مفيداً، إلا أن السرد الذي ينقل التداعيات الواقعية لمسألة ما، يمكن أن يكون أكثر مباشرة وتأثيراً على القارئ.

عندما تكون لديك نقاط بيانات قليلة جداً

قيل أن "الرقم بمفرده لا يعني شيئاً." وهناك رد شائع من محرري الأخبار على الإحصائيات المقتبسة وهو: "بالمقارنة مع ماذا؟" هل يرتفع المعدل أم ينخفض، وما هو المعدل العادي؟

عندما يكون لديك القليل من المتغيرات في بياناتك، بدون أي اتجاه واضح أو استنتاج

أحياناً تخطط لبياناتك في إكسل أو تطبيق بياني مشابه وتكتشف أن المعلومات مضطربة، أي بها الكثير من التقلبات أو دون اتجاه ثابت نسبياً. هل سترفع الخط من الصفر إلى ما دون الحد الأدنى بقليل حتى تعطيه شكلاً ما؟ لا! إذ يبدو الأمر كأن لديك بيانات غامضة وتحتاج إلى المزيد من التنقيب والتحليل.

عندما تكون الخريطة ليست فعلياً خريطة

أحياناً لا يكون العنصر المكاني مهماً أو جذاباً، أو قد يشتت الانتباه من الاتجاهات الرقمية الأكثر صلة، مثل التغير عبر الزمن أو إظهار التشابه بين مناطق غير متجاورة.

لا تنس الجداول

إذا كانت لديك نقاط بيانات قليلة نسبياً ولكن لديك معلومات قد تكون مفيدة لبعض قرائك، فكر في وضع البيانات في جداول، فهي نظيفة وتسهل قراءتها ولا تخلق توقعات غير واقعية للخبر، بل ويمكن للجداول أن تكون نموذجاً فعّالاً وملائماً جداً للمعلومات الأساسية.

—— جيف ماكغي، جامعة ستانفورد

تختلف القصة باختلاف الرسوم البيانية

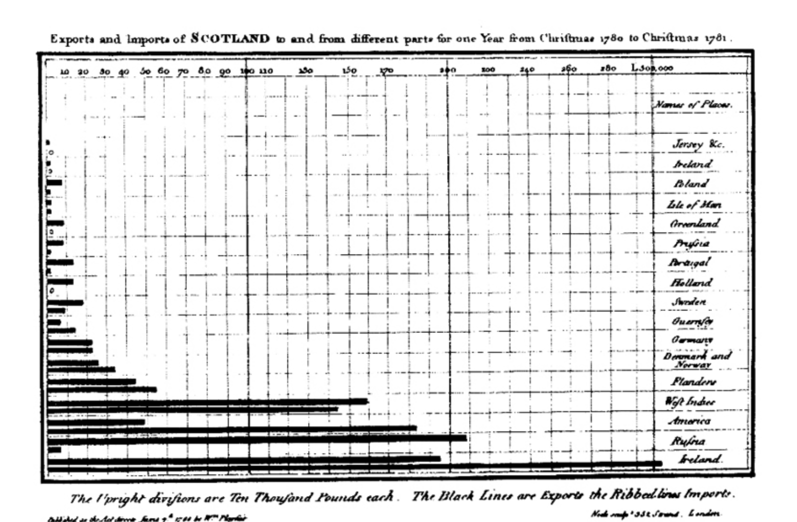
في هذا العالم الرقمي، ومع الأمل في تحقيق تجارب ثلاثية الأبعاد، نميل إلى نسيان أننا لمدة طويلة جداً لم يكن لدينا سوى الحبر والورق. نحن الآن نعتبر هذا الوسط الثابت والمسطح عنصراً من الدرجة الثانية، ولكن في الواقع وعلى مدى مئات السنوات التي كتبنا وطبعنا خلالها، تمكنا من تحقيق ثروة هائلة من المعرفة والممارسات لوضع البيانات على الصفحات. وعلى الرغم من أن الرسوم البيانية التفاعلية والتوضيحات المرئية للبيانات والإنفوغرافيكس هي الأكثر رواجاً، فإنها تتخلى عن أفضل الممارسات التي تعلمناها. فقط عندما نرجع إلى تاريخ إنجازات الرسوم البيانية ستتمكن من فهم هذا البنك المعرفي لنعبر به إلى وسائط جديدة.

إن بعض أشهر الرسوم البيانية نتجت عن الحاجة إلى شرح جداول البيانات المكثفة بشكل أفضل. كان ويليام بلايفير اسكوتلندي متعدد اللغات وعاش بين أواخر القرن الثامن عشر وأوائل القرن التاسع عشر. وقد قدم بمفرده للعالم العديد من الرسوم البيانية ذاتها التي ما زلنا نستخدمها حتى اليوم. وفي كتابه الذي نُشر عام 1786 تحت عنوان 'الأطلس التجاري والسياسي'، قدم بلايفير الرسم البياني الشريطي ليوضح حجم الاستيراد والتصدير في اسكوتلندة بشكل جديد ومرئي.

قام بلايفير بعد ذلك بنشر الرسم البياني الدائري في كتابه الذي نُشر عام

1801 تحت عنوان ' الملخص الإحصائي'. نتجت الحاجة إلى تلك الأشكال الجديدة للرسوم البيانية عن التجارة، ولكن مع مرور الوقت، ظهرت أشكال أخرى تم استخدامها لإنقاذ حياة الناس. في عام 1854، رسم جون سنو 'خريطة انتشار الكوليرا في لندن' وهي خريطة مشهورة، حيث وضع مربعاً أسود صغيراً فوق كل عنوان تم التبليغ فيه عن حالة عدوى. وبمرور الوقت، كان يمكن رؤية أماكن الكثافة الواضحة لانتشار المرض وتم اتخاذ الخطوات للتصدي للمشكلة.

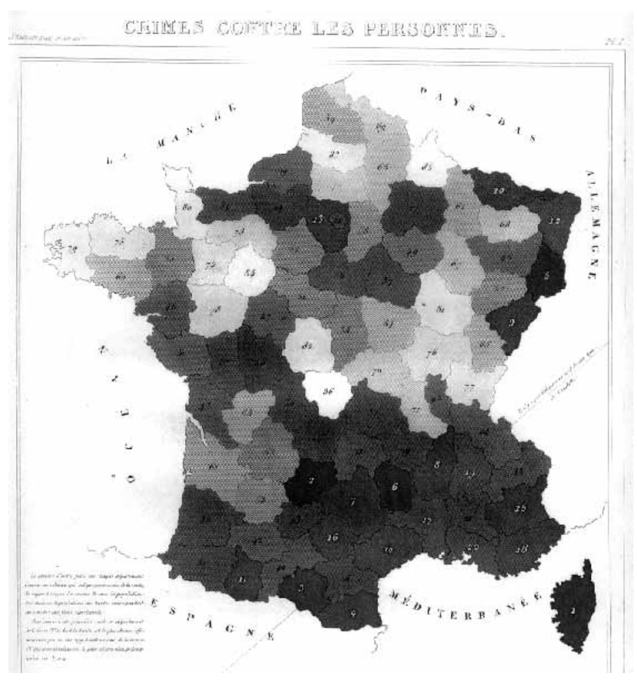
مع مرور الوقت، تجرأ ممارسو تلك الرسوم البيانية أكثر وأكثر وقاموا بالمزيد من التجارب، دافعين بها إلى ما نراه اليوم. كان أندري–ميشيل غيري أول من نشر فكرة الخريطة ذات المناطق المختلفة الملونة حسب بعض المتغيرات. في عام 1829، رسم أول خريطة تظليل نسبي Choropleth Map عبر تظليل مناطق معينة في فرنسا لتمثيل معدل الجريمة. نرى اليوم مثل تلك الخرائط لعرض الدوائر الانتخابية، ومن أدلى بصوته لمن، وتوزيع الثروات، ومتغيرات أخرى عديدة مرتبطة جغرافياً. تبدو كأنها فكرة بسيطة للغاية، ولكن حتى اليوم، تصعب إجادتها وفهمها إن لم يتم استخدامها بحكمة.



نموذج 6–11. رسم بياني شريطي قديم (ويليام بلايفير)



نموذج 6–12. خريطة انتشار الكوليرا في لندن (جون سنو)

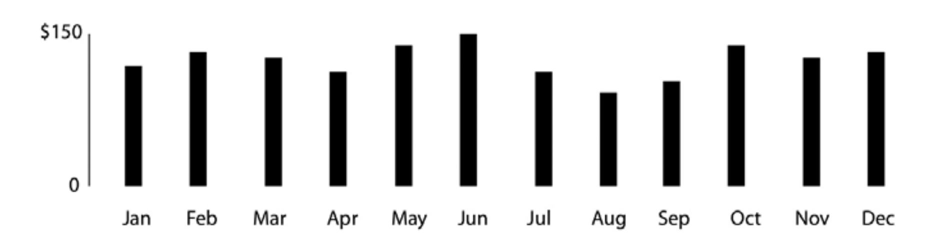


نموذج 6–13. خرائط التظليل النسبي (كوروبلِث) Choropleth map لعرض معدلات الجريمة في فرنسا (أندري–ميشيل– غيري)

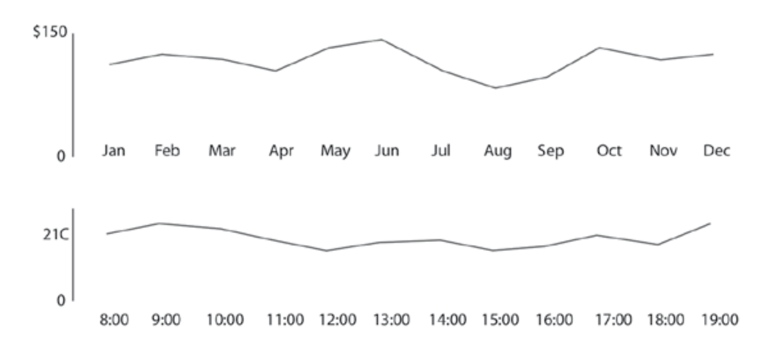
هناك العديد من الأدوات التي يحتاج الصحفي الجيد إلى فهمها ووضعها ضمن أدواته لبناء التوضيحات المرئية. وبدلاً من البدء بالجزء الصعب، من المهم الحصول على تمهيد ومعرفة أساسية توفرها الرسوم البيانية. فكل ما تقوم به يجب أن ينبع من سلسلة رسوم بيانية مبسطة. إذا تمكنت من إجادة الأساسيات، يمكنك أن تنتقل إلى بناء توضيحات مرئية أكثر تعقيداً مكونة من تلك الوحدات الأساسية.

هناك نوعان من أكثر أنواع الرسوم البيانية أساسية وهما الرسم البياني العمودي والرسم البياني الخطي. وعلى الرغم من أنهما متشابهان جداً من حيث حالات استخدامهما، إلا أنهما قد يختلفان كثيراً في معانيهما. فإذا أوضحنا على سبيل المثال مبيعات الشركة في كل شهر من شهور السنة. سنحصل على 12 عمودا يمثل كل منها حجم الإيرادات في كل شهر (نموذج 6–14).

لننظر لماذا يجب على هذا الرسم البياني أن يكون عمودياً وليس خطياً. تعتبر الرسوم البيانية الخطية مثالية للبيانات المتواصلة. أما بالنسبة لأرقام مبيعاتنا، فهي مجموع الأشهر، وليست متواصلة. من خلال العمود نعلم أن الشركة كسبت 100 دولار في يناير وكسبت 120 دولار في فبراير. أما إذا جعلنا هذا الرسم البياني خطياً، فسيمثل مبلغا 100 و120 دولاراً في الأول من الشهرين، ولكن مع الرسم البياني الخطي سنقدّر أن الشركة حصلت على 110 دولار في الخامس عشر من الشهر، وهذا ليس صحيحاً. يُستخدم العمود لتمثيل وحدات القياس المنفصلة، بينما يُستخدم الخط لتمثيل القيمة المتواصلة مثل درجات الحرارة.



نموذج 6–14. رسم بياني عمودي بسيط: مفيد في تمثيل المعلومات المنفصلة

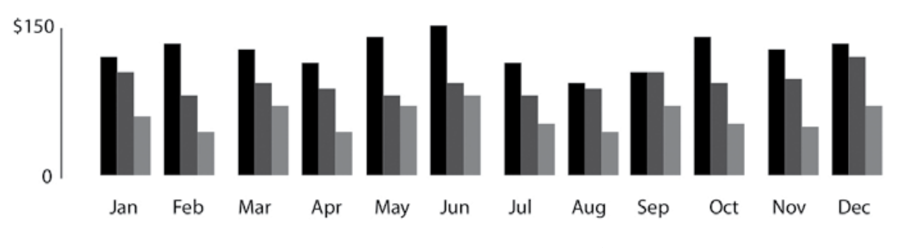


نموذج 6–15. رسوم بيانية خطية بسيطة: مفيدة في تمثيل المعلومات المتواصلة

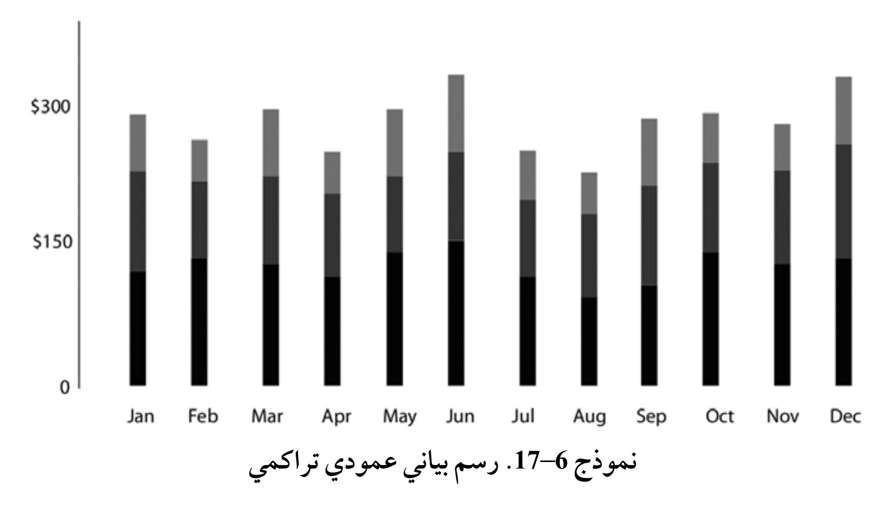
يمكننا أن نرى أن درجة الحرارة كانت 20 درجة مئوية في الثامنة صباحاً بينما ارتفعت إلى 22 درجة في التاسعة صباحاً. وإذا نظرنا إلى الخط لتقدير درجة الحرارة في الثامنة والنصف سنقول أنها كانت 21 درجة مئوية، وهو تقدير صحيح حيث أن درجات الحرارة متواصلة وكل نقطة لا تمثل مجموع الأرقام الأخرى، بل تمثل القيمة المحددة في تلك اللحظة أو تقديراً يتراوح بين قياسين محددين.

لدى كل من العمود والخط تغيرات تراكمية، وهذه أداة ممتازة لسرد الأخبار ويمكنها العمل بطرق مختلفة. دعونا نرى على سبيل المثال شركة لديها ثلاثة مقرات.

لكل شهر لدينا ثلاثة أعمدة، واحد لكل مقر، بإجمالي 36 شريطاً للعام. وعندما نضع تلك الأعمدة جنباً إلى جنب، يمكننا أن نرى في الحال أي المقرات حصلت على عوائد أكبر وفى أي من الأشهر. وهذا خبر دقيق ومثير للاهتمام، ولكن هناك خبراً آخر متضمناً في البيانات ذاتها. إذا جمعنا الأعمدة ليكون لدينا عمود واحد فقط لكل شهر، لن نتمكن من رؤية أي المقرات تحصل على أكبر عائد، ولكننا سنتمكن من معرفة الأشهر التي تحقق فيها الشركة عوائد مالية.



نموذج 6–16. رسم بياني عمودي مجمع



نموذج 6–17. رسم بياني عمودي تراكمي

كلتا الوسيلتين تقدمان عرضاً دقيقاً لنفس المعلومات، ولكنهما تشكلان خبرين مختلفين باستخدام نفس البيانات المبدئية. وبالنسبة للصحفي، فإن أهم جانب من العمل مع البيانات هو أولاً اختيار الخبر الذي تود سرده. هل هو أي الأشهر يحمل عائداً أكبر؟ أم هو أي المقرات حقق عائدات أكبر؟ هذا مجرد مثال بسيط، ولكنه بالفعل هو جوهر العمل الصحفي المتمثل في طرح السؤال الصحيح قبل أن تبتعد كثيراً. سيرشدك الخبر إلى اختيار نوع التوضيح المرئي المناسب.

الرسمان البيانيان العمودي والخطي هما حقاً العنصران الأساسيان لأي صحفي بيانات. ومنهما يمكنك التوسع إلى المدرج الإحصائي والرسم البياني بالأفق والخطوط البيانية والرسم البياني التدفقي وأشكال أخرى تتشارك جميعها في الخصائص وتناسب كل منها حالات متباينة بدرجات متفاوتة، بما في ذلك كم البيانات أو مصادرها وموقع الرسم البياني بالنسبة للنص المكتوب.

الخريطة هي إحدى أكثر الخصائص البيانية استخداماً في الصحافة. فالوقت والمقدار والجغرافيا كلها عناصر شائعة في الخرائط. نريد دائماً أن نعرف مقدار شيء ما في منطقة ما بالمقارنة مع منطقة أخرى أو كيف تتدفق البيانات من منطقة إلى أخرى. ومن الأدوات المفيدة جداً التي يجب أن تكون لديك في مجموعة مهاراتك عند التعامل مع التوضيح المرئي في الصحافة مخططات التدفق وخرائط التظليل النسبي (Choropleth). ومن المهم للغاية أن تعرف كيفية رسم خريطة مرمزة باستخدام الألوان بدون تشويه أو تضليل للقراء. عادة ما تكون الخرائط

السياسية مرمزة بالألوان وعادة ما تكون ثنائية الألوان يرمز أحد اللونين إلى طرف واللون الآخر إلى الطرف المقابل، وهذا أمر لا يعكس الواقع بدقة، فليس من الضروري أن تكون الألوان ثنائية، فيمكن استخدام التدرجات اللونية بحرص حسب الدلالة. ويُشكل فهم الخرائط جزءاً كبيراً من الصحافة، فالخرائط تجيب بسهولة على الجزء الخاص بـ 'المكان' من الأسئلة.

بمجرد إجادتك للأنواع الأساسية من الرسوم البيانية يمكنك أن تبدأ في بناء توضيحات مرئية أكثر تميزاً. إذا لم تفهم الأساسيات وتستوعبها جيدا، ستبني على أساس هش. بنفس الأسلوب المماثل للطريقة التي تتعلم بها كيف تكون كاتباً جيداً – مثل جعل الجمل قصيرة وأخذ الجمهور في الحسبان وعدم تعقيد الأمور عمدا لتجعل نفسك تبدو أكثر ذكاءً ونقل المعنى مبسطا إلى القارئ – عليك ألا تبالغ في البيانات أيضاً، فالبدء بالأمور الصغيرة يُعد أكثر الطرق فعالية لسرد الخبر، مع البناء تدريجيا وببطء فقط عند الحاجة لذلك.

'إن الكتابة الجيدة هي الكتابة الموجزة. يجب ألا تحتوي الجملة على كلمات لا ضرورة لها، كما يجب ألا تحتوي الفقرة على جُمل غير ضرورية، وهذا لنفس السبب الذي يحتم على الرسم ألا يحتوي على خطوط غير ضرورية وعلى الآلات ألا تحتوي على أجزاء غير ضرورية. وهذا لا يتطلب أن يجعل الكاتب جُمله قصيرة أو يتجنب كافة التفاصيل ويعرض مواضيعه بشكل مُختصر، بل يتطلب منه أن يكون لكل كلمة يكتبها معنى.'

— ويليام سترانك جونيور (عناصر الأسلوب – 1918)

لا ضير من إهمال إجزاء شتى من البيانات المتاحة لديك في الخبر. لست بحاجة لطلب الإذن كي توجز، بل يجب أن تكون تلك هي القاعدة.

— بريان سودا، (optional.is)

التوضيح المرئي للبيانات، كيف تنفذه بنفسك: أفضل أدواتنا

ما هي أدوات التوضيح المرئي للبيانات المتاحة مجاناً على شبكة الإنترنت، والتي يسهل استخدامها؟ هنا في مدونة البيانات ومتجر البيانات (.http://www

guardian.co.uk/data) نحاول أن نفعل ما بوسعنا باستخدام الخيارات الجيدة المجانية على شبكة الإنترنت.

قد يبدو هذا مخادعاً، إذ يمكننا بالطبع الوصول إلى العالمين بفرق الرسوم البيانية المذهلة والفرق التفاعلية في الغارديان المختصين بتلك الأعمال عندما يتاح لدينا مزيدا من الوقت - مثل تلك الخريطة المتعلقة بالإنفاق العام (http://bit.ly/guardian-spending) التي رسمناها باستخدام Adobe Illustrator أو هذا الرسم التفاعلي لأحداث الشغب على موقع تويتر (http://bit.ly/guardian-riots).

ولكن بالنسبة لعملنا اليومي، عادة ما نستخدم أدوات يمكن للجميع استخدامها ونبني رسوماً بيانية يمكن للجميع أيضاً بناؤها.

إذن فماذا نستخدم؟

جداول جوجل فيوجن Google Fusion Tables

لقد أصبحت قاعدة البيانات وأداة رسم الخرائط تلك أساسية لنا لعمل خرائط سريعة وتفصيلية، وخاصة تلك التي تحتاج إلى التكبير. تحصل على كل خرائط جوجل عالية الدقة ولكن يمكنها فتح الكثير من البيانات، مثل 100 ميجابايت من ملفات CSV مثلاً. عندما تجرب جداول فيوجن للمرة الأولى، قد تبدو صعبة قليلاً، ولكن استمر في استخدامها. لقد استخدمناها لوضع خرائط مثل خريطة العراق في نموذج 6–18 وخرائط حدودية مثل نموذج 6–19 الخاصة بالتشرد.

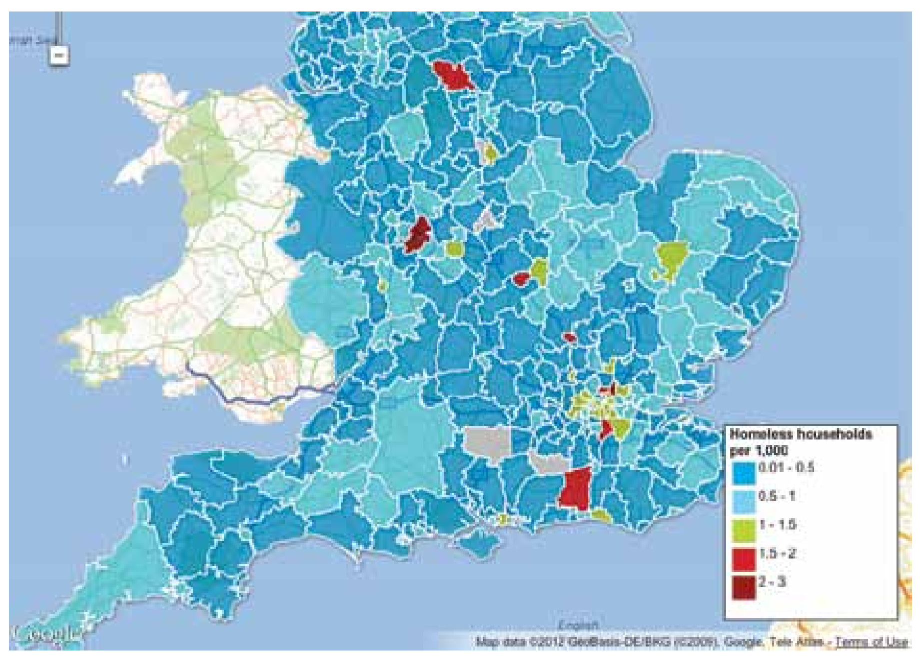
إن أهم الميزات هي المرونة – يمكنك تحميل ملف KML عن خريطة الحدود الإقليمية مثلاً – ثم دمجه مع جدول بيانات. كما أنها ستحصل على وصلة إنترفيس مستخدم جديدة ستسهل من استخدامها.

ليس من الضروري أن تكون مبرمجاً حتى تصنع خريطة – فأداة فيوجن للطبقات (http://bit.ly/fusion–layers) تتيح لك جمع خرائط مختلفة معاً أو خلق خيارات للبحث والفرز يمكنك أن تضيفها إلى مدونة أو موقع إلكتروني.

وهذا الدرس الممتاز من كاثرين هيرلي التي تعمل في جوجل (http://bit.ly/fusiontables–tutorial) يعتبر مكانا ملائماً للبدء.



نموذج 6–18. سجلات الحرب من ويكيليكس (الغارديان)



نموذج 6–19. خريطة التشرد التفاعلية (الغارديان)

نصيحة: استخدم shpescape (http://www.shpescape.com/) لتحويل ملفات shp الرسمية إلى جداول فيوجن لتتمكن من استخدامها. احذر أيضاً من الخرائط المعقدة، فلا يمكن لفيوجن التعامل مع أكثر من مليون نقطة في الخلية الواحدة.

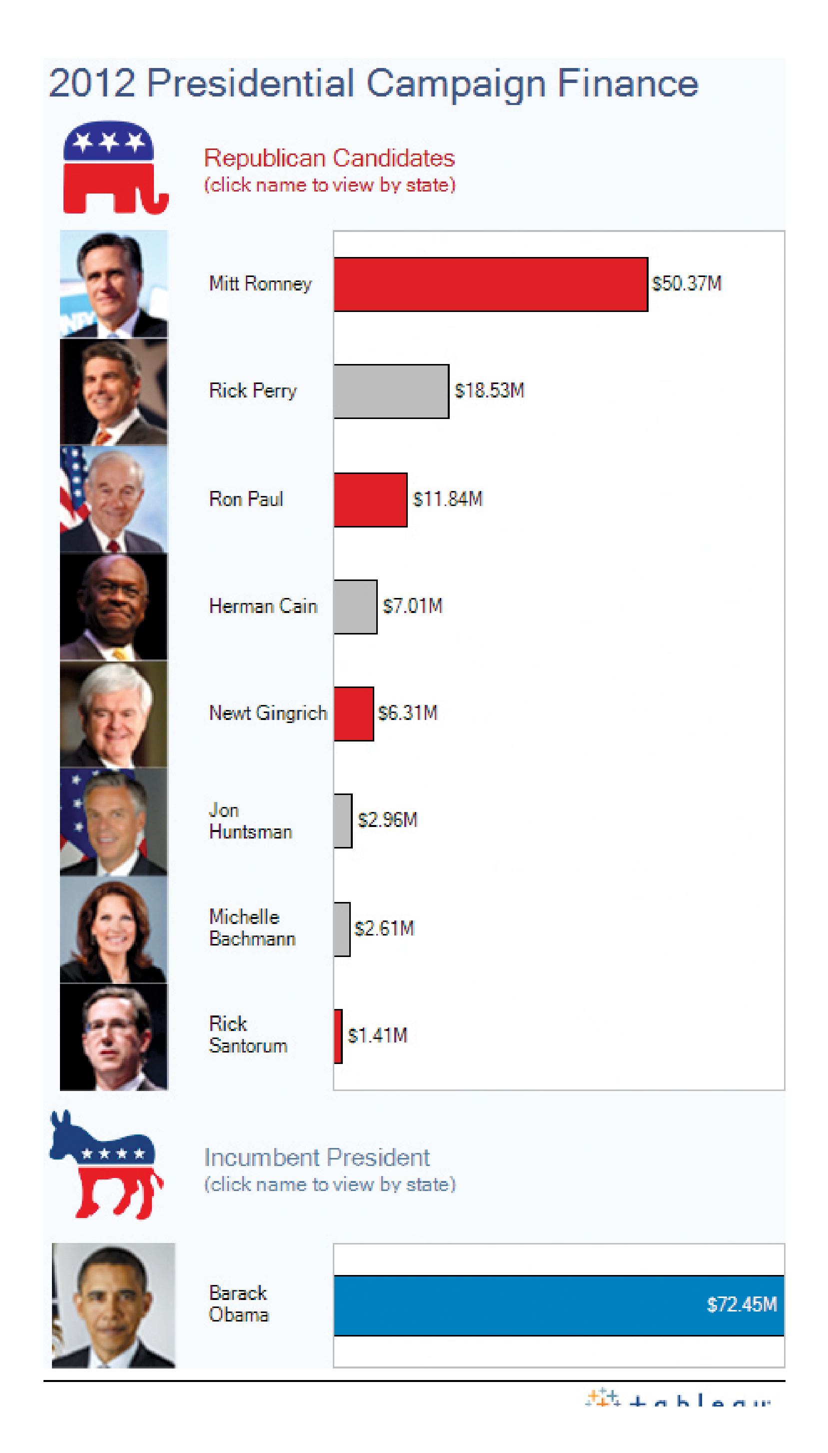
برنامج Tableau Public

إذا لم تكن في حاجة إلى المساحة غير المحدودة في الإصدار المحترف، فبرنامج Tableau Public (http://www.tableausoftware.com/public) المجاني يمكنّك من أن تبني توضيحات مرئية معقدة بما يصل إلى 100,000 صف ببساطة وسهولة. نحن نستخدمه عندما نريد جمع أنواع مختلفة من الرسوم البيانية، مثلما فعلنا في تلك الخريطة لأعلى معدلات الضرائب في العالم (http://bit.ly/guardian-top-tax) والتي تحتوي أيضاً على رسم بياني عمودي.

أو يمكنك حتى استخدام البرنامج كمستكشف للبيانات، وهو ما فعلناه في نموذج 6-20 مع بيانات إنفاق الانتخابات الفيدرالية الأمريكية (http://bit.ly/guardianelections-us) على الرغم من أن المساحة قد نفدت منا في الإصدار المجاني العام. يحتاج البرنامج أيضاً أن تكون البيانات في صيغ محددة حتى تحصل على الاستفادة القصوى منها. ولكن إذا تخطيت ذلك، سيكون لديك شيء سهل ويعمل بشكل جيد. وقد قامت صحيفة لا ناسيون الأرجنتينية مثلاً ببناء عملية صحافة البيانات بأكملها على برنامج تابلو (http://blogs.lanacion.com.ar/data/).

يحوى برنامج تابلو بعض الدروس الجيدة على الإنترنت لتبدأ بها (http://www.tableausoftware.com/learn/training).

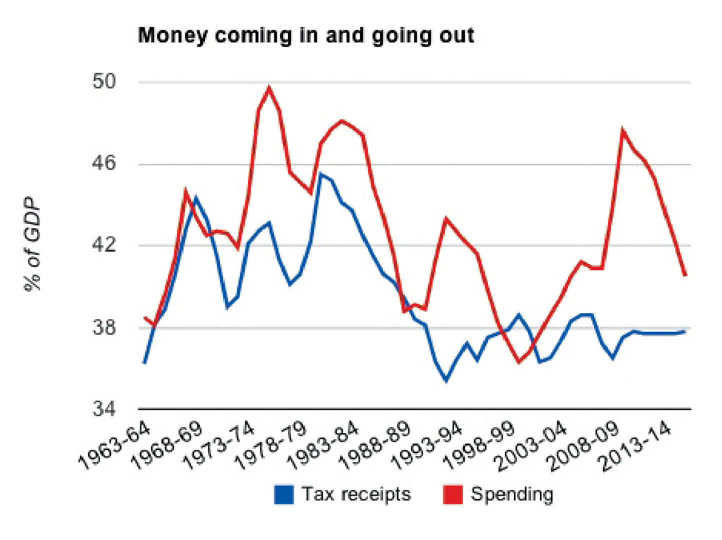
نصيحة: تم تصميم برنامج تابلو لأجهزة الكمبيوتر PC، ولكن هناك نسخة لأجهزة Mac في طور التجهيز. استخدم مرايا مثل برنامج parallels ليعمل على جهازك.



نموذج 6–20. تمويل الحملات الانتخابية الرئاسية عام 2012 (الغارديان)

جداول جوجل البيانية

يمكنك الحصول على هذه الأداة من الرابط التالي (http://www.google.com/google-d-s/spreadsheets/).



نموذج 6–21. الإنفاق والضرائب لدى الحكومة البريطانية (الغارديان)

بعد التعامل مع شيء سهل (مثل الرسم البياني العمودي أو الخطي أو الدائري)، ستجد أن جداول جوجل البيانية (التي تصنعها من الجزء الخاص بالوثائق في حساب جوجل الخاص بك) يمكنها أن ترسم رسوما بيانية جيدة جداً – بما في ذلك الفقاعات المتحركة التي استخدمها هانز روزلينغ في موقعه الإلكتروني Gapminder (http://www.gapminder.org/). وعلى خلاف وصلة إنترفيس برمجة التطبيقات لدى الرسوم البيانية (http://code.google.com/apis/chart/)، لا تقلق إزاء الترميز (الكود)، فالأمر مشابه لصنع الرسم البياني باستخدام إكسل حيث تظلل البيانات وتنقر على واجهة الأدوات. تستحق خيارات التخصيص أيضاً أن تستكشفها، حيث يمكنك تغيير الألوان والعناوين والمقاييس. كما أن التصميم محايد مما يجعله مفيداً في الرسوم البيانية الصغيرة. وللرسوم البيانية الخطية خيارات جيدة أيضاً، بما فيها خيارات الحاشية.

نصيحة: أمض بعض الوقت في تصفح خيارات التخصيص للرسم البياني،

وسيمكنك أن تصنع لوحة الألوان الخاصة بك.

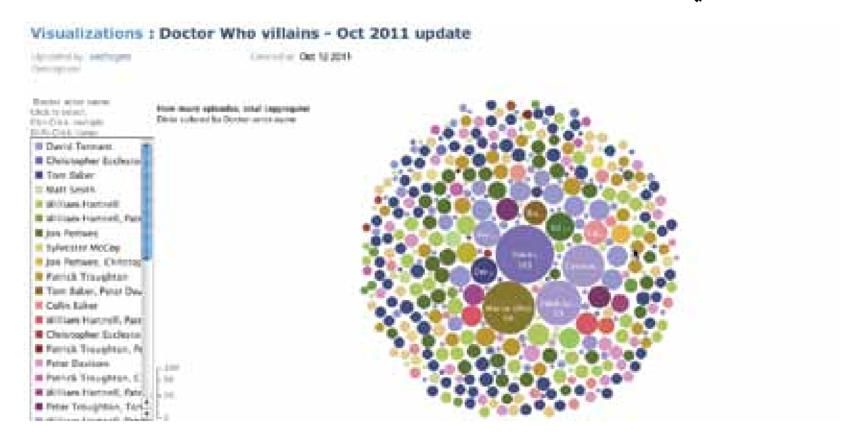
شركة Datamarket

إن شركة Datamarket المعروفة أكثر بكونها شركة موردة للبيانات هي فعليا أداة أنيقة لتوضيح الأرقام مرئياً أيضاً. يمكنك تحميل مجموعة بياناتك أو استخدام بعض من مجموعات البيانات العديدة التي يقدمونها، ولكن الخيارات تتحسن إذا حصلت على حساب المحترفين.

نصيحة: تعمل أداة Datamarket بشكل أفضل مع البيانات المتسلسلة زمنيًا، ولكن تحقق أيضاً من مجموعة بياناتهم الشاملة.

موقع Many Eyes

إذا كان هناك موقع واحد يحتاج إلى بعض الرعاية فهو موقع Many Eyes التابع لـ IBM (http://ibm.co/ibm-manyeyes). فعندما تم إطلاقه بواسطة فيرناندا فيغاس (http://fernandaviegas.com/) ومارتن واتنبيرغ (http://www.bewitched.com/)، كان تجربة فريدة في السماح للناس بتحميل مجموعات البيانات ببساطة وتوضيحها مرئياً. والآن بعد انتقال مؤسسي الموقع إلى العمل في جوجل، يشعر الموقع بنقص في الرعاية بسبب لوحة الألوان المكتومة ولم يشهد تغيرات عديدة في التوضيحات المرئية منذ فترة.



نموذج 6-22. رسم يوضح الشخصيات الشريرة في مسلسل 'دكتور هو' Doctor Who (الغارديان) (http://bit.ly/guardian-dr-who)

موقع Color Brewer

إن موقع Color Brewer (http://colorbrewer2.org) الذي تم تصميمه في الأساس بدعم مالي فيدرالي وقامت ولاية بنسلفانيا بتطويره، ليس أداة توضيح مرئي تحديداً، بل هو أصلاً لاختيار ألوان الخرائط. يمكنك اختيار اللون الأساسي والحصول على (الكود) للوحة الألوان بأكملها.

والمزيد

إن لم يناسبك أي من ذلك، تحقق من هذا الرسم في DailyTekk (http://bit.ly/dailytekk-infographic) الذي لديه المزيد من الخيارات. إن ما ذكرناه أعلاه ليس الأدوات الوحيدة، بل فقط الأدوات التي نستخدمها نحن على نحو متكرر. هناك العديد من الأدوات الأخرى، بما فيها:

- Chartsbin (http://chartsbin.com/) وهي أداة لصنع خرائط للعالم قابلة للنقر عليها.
- iCharts (http://www.icharts.net/) التي تتخصص في واجهات الأدوات للرسوم البيانية الصغيرة.
- Geocommons (http://geocommons.com/) التي تشارك بيانات الحدود لصنع الخرائط العالمية والمحلية.
- وهناك أيضاً موقع piktochart.com (http://piktochart.com/) الذي يوفر قوالب لتلك التوضيحات المرئية ذات النصوص/الأرقام الشائعة حالياً.

— سايمون روجرز، الغارديان

كيف نقدم البيانات في صحيفة فيردنس غانج Verdens Gang

تعني صحافة الأخبار تقديم معلومات جديدة للقارئ بأسرع ما يمكن. قد تكون أسرع وسيلة الفيديو أو الصورة أو النص أو الرسم البياني أو الجدول أو حتى مزيجاً من تلك الوسائل. بالنسبة للتوضيحات المرئية، يجب أن تكون لنفس الغرض: المعلومات السريعة. تتيح أدوات البيانات الجديدة للصحافيين أن يجدوا أخباراً لم يكن بإمكانهم العثور عليها بدون تلك الأدوات، كما تتيح لهم تقديم

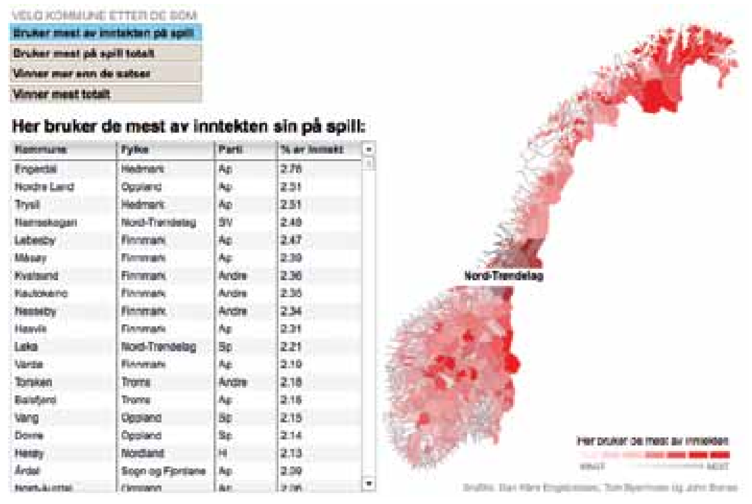
الأخبار بطرق جديدة. فيما يلي بعض الأمثلة التي توضح كيف نقدم البيانات في أكثر الصحف قراءة في النرويج، صحيفة فيردنس غانج (VG).

الأرقام

وضع هذا الخبر (http://bit.ly/vg–lotto) بناء على بيانات من مكتب الإحصاءات النرويجي وبيانات دافعي الضرائب وبيانات من مكاتب احتكار اليانصيب الوطني. في هذا الرسم البياني التفاعلي، يستطيع القارئ أن يجد أنواعاً مختلفة من المعلومات من كل مقاطعة وبلدية نرويجية. ويُظهر الجدول الفعلي نسبة الدخل المستخدمة في الألعاب، وقد تم بناؤه باستخدام برامج أكسيس Access وإكسل Excel وMySql وفلاش Flash.

الشبكات

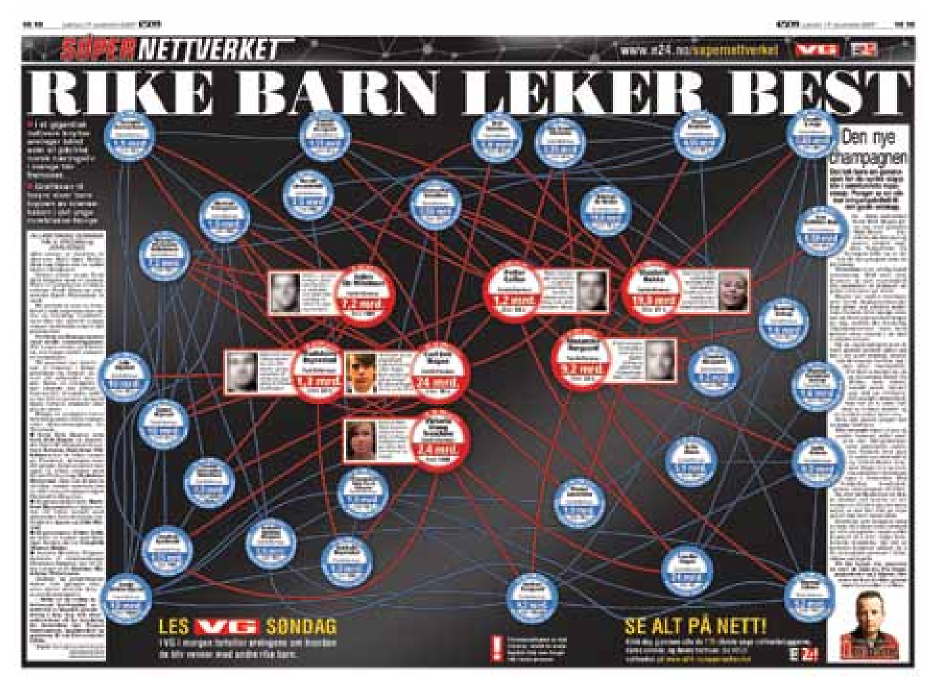
استخدمنا تحليلات الشبكات الاجتماعية لنحلل العلاقات بين 157 ابنا وابنة لأغنى أغنياء النرويج. وقد أظهر تحليلنا أن ورثة أغنياء النرويج قد ورثوا أيضاً شبكات أعمال آبائهم. كان هناك إجمالاً ما يزيد عن 26,000 صلة، وقد أكملنا الرسوم البيانية كلها يدوياً باستخدام برنامج فوتوشوب Photoshop. واستخدمنا أيضاً برامج أكسيس Access وإكسل Excel ونوتباد Notepad وأداة تحليل الشبكات الاجتماعية Ucinet.



Kommune	Fylke	Parti	% av inntekt
Engerdal	Hedmark	Ap	2.76
Nordre Land	Oppland	Ap	2.51
Trysil	Hedmark	Ap	2.51
Namsskogan	Nord-Trøndelag	SV	2.49
Lebesby	Finnmark	Ap	2.47
Måsøy	Finnmark	Ap	2.39
Kvalsund	Finnmark	Andre	2.36
Kautokeino	Finnmark	Andre	2.35
Nesseby	Finnmark	Andre	2.34
Hasvik	Finnmark	Ap	2.31
Leka	Nord-Trøndelag	Sp	2.21
Vardø	Finnmark	Ap	2.19
Torsken	Troms	Andre	2.18
Balsfjord	Troms	Ap	2.16
Vang	Oppland	Sp	2.15
Dovre	Oppland	Sp	2.14
Herøy	Nordland	H	2.13
Årdal	Sogn og Fjordane	Ap	2.09
[illegible]	[illegible]	[illegible]	[illegible]

نموذج 6–23. تخطيط بيانات دافعي الضرائب وبيانات اليانصيب (صحيفة فيردنس غانج)

في تلك الخريطة الملونة المتحركة والمجتمعة رسم بياني شريطي بسيط (http://bit.ly/vg–heatmap)، يمكنك أن ترى جرائم تحدث على خريطة لقلب مدينة أوسلو ساعة بساعة وخلال عطلات نهاية الأسبوع على مدى عدة أشهر. وفي نفس الخريطة الملونة المتحركة، يمكنك أن ترى عدد رجال الشرطة الذين يعملون في نفس الوقت. وعند حدوث الجريمة، سترى عدد رجال الشرطة في أسفل الخريطة. وقد تم بناؤها باستخدام ArcView مع Spatial Analyst.


SUPER NETTVERKET
www.e24.no/supernettverket
RIKE BARN LEKER BEST
Den nye champagnen
LES VG SØNDAG
SE ALT PÅ NETT!

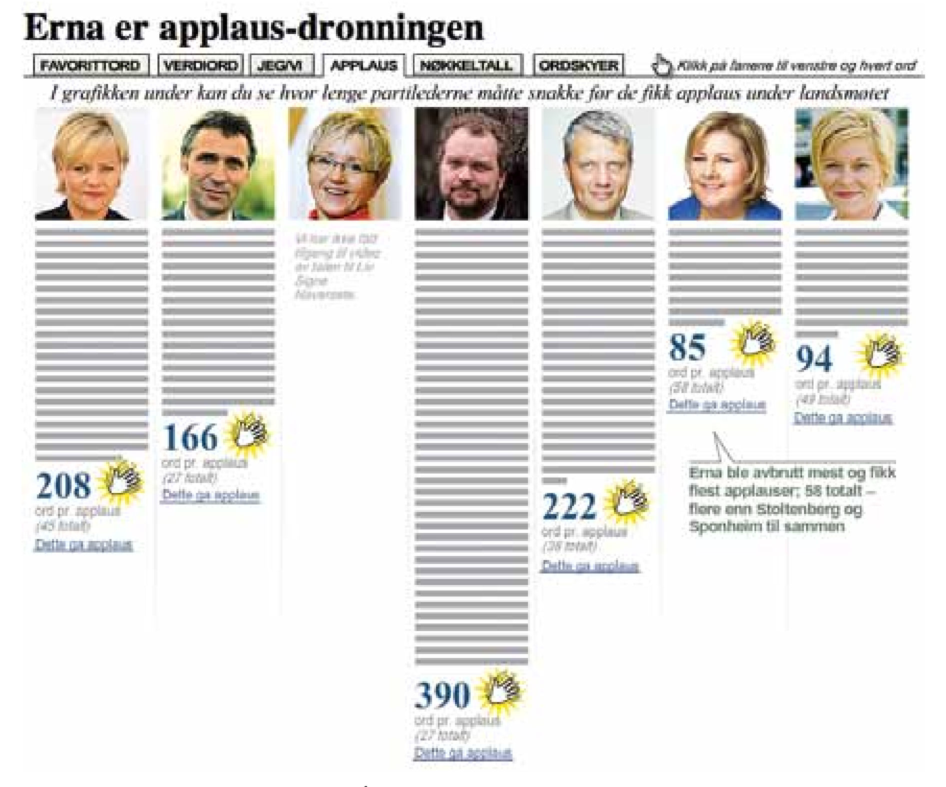
نموذج 6–24. الطيور على أشكالها تقع: الأغنياء وأبناؤهم (صحيفة فيردنس غانج)



نموذج 6–25. خريطة ملونة متحركة (صحيفة فيردنس غانج)

التنقيب في النصوص

لهذا التوضيح المرئي (http://bit.ly/vg-vis) قمنا بالتنقيب في نصوص خطابات رؤساء الأحزاب النرويجية السبعة أثناء مؤتمراتهم. تم تحليل كل الخطابات مما زود بعض المقاطع الخبرية بزوايا مختلفة. كما تم ربط كل خبر بالرسم البياني مما أتاح للقراء استكشاف لغة السياسيين ودراستها. وقد بنينا ذلك باستخدام برامج أكسيس Access وإكسل Excel وفلاش Flash وإلستريتور Il-lustrator. وإذا بنينا هذا الرسم البياني التفاعلي عام 2012، كنا سنستخدم برنامج جافاسكريبت Javascript.



نموذج 6–26. التنقيب في نصوص خطابات رؤساء الأحزاب (صحيفة فيردنس غانج)

الملاحظات الختامية

متى نحتاج إلى توضيح الخبر مرئياً؟ في أغلب الأحيان لا نحتاج إلى عمل ذلك، ولكن أحياناً نريد أن نقوم بذلك لنساعد قرائنا. عادة ما تحتاج الأخبار التي تحتوي على كم هائل من البيانات إلى توضيح مرئي. ولكن علينا أن نكون حريصين للغاية عند اختيار نوع البيانات التي ننوي تقديمها. فنحن نعلم كل الأمور عندما نعد التقرير، ولكن هل يحتاج القارئ فعلاً إلى معرفة كل تلك الأمور التي عرفناها نحن؟ ربما يكفي الجدول أو الرسم البياني البسيط الذي يوضح تطوراً ما من العام (أ) إلى العام (ب). وعند العمل مع صحافة البيانات، لا يكون الهدف بالضرورة تقديم كمٍ هائلٍ من البيانات، فالأمر يتعلق بالصحافة!

كان هناك اتجاهٌ واضح في العامين أو الثلاثة أعوام الأخيرة نحو صنع رسوم

بيانية تفاعلية وجداول تتيح للقارئ التعمّق في شتى المواضيع. فالتوضيح المرئي الجيد أصبح كالصورة الجيدة، من حيث القدرة على فهم موضوعها بمجرد النظر إليها للحظة قصيرة. وكلما نظرت أكثر إلى ما هو مرئي رأيت أكثر وازدادت ملاحظاتك. ويكون التوضيح المرئي سيئاً عندما لا يعلم القارئ أين يبدأ وأين يتوقف، وأيضاً عندما يكون التوضيح المرئي محشواً بالتفاصيل. في تلك الحالة، ربما يكون النص هو الأفضل.

— جون بونز، صحيفة فيردنس جانج

البيانات العامة تصل إلى الشبكات الاجتماعية

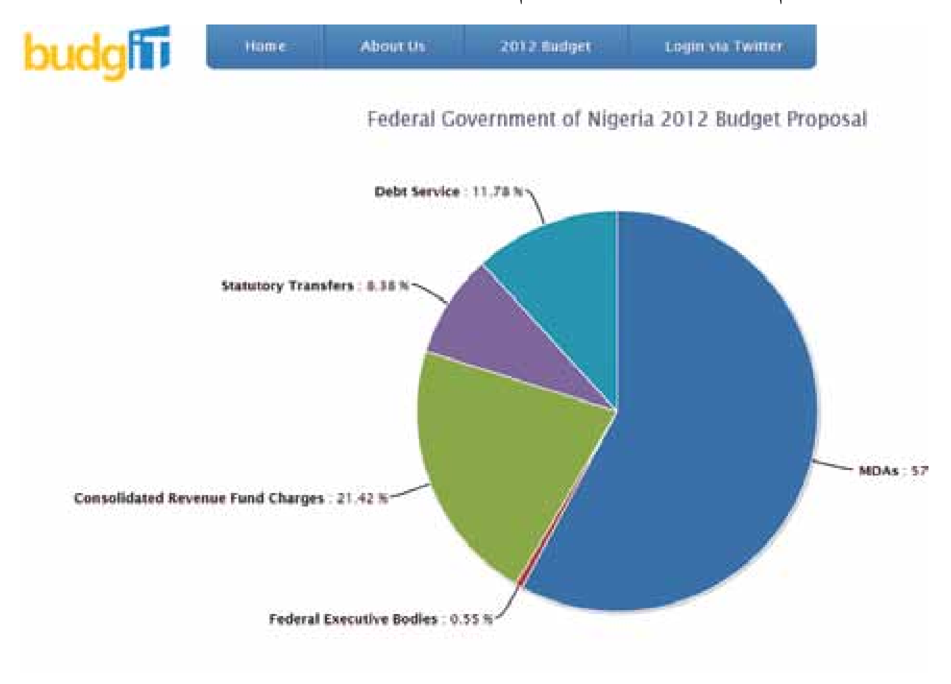
البيانات لا تقدّر بثمن. والوصول إليها يمكننا من تسليط الضوء على القضايا بشكل يُحقق نتائج ملموسة. ومع ذلك، يمكن لسوء التعامل مع البيانات أن يضع الحقائق في هيكل غامض لا ينقل شيئاً. وإذا لم تعزز المناقشات أو توفر فهماً ضمن سياق، فقد تكون البيانات ذات فائدة محدودة.

عادت نيجيريا إلى الديمقراطية عام 1999 بعد سنوات طويلة من الحكم العسكري. وقد اعتبر التحقيق في الحقائق وراء البيانات إهانة للسلطة ومحاولة للتشكيك في سمعة العسكر الملطخة بالفساد. وقد منع 'قانون الأسرار الرسمية' الموظفين الحكوميين من نشر المعلومات الحكومية. وحتى بعد مرور ثلاثة عشر عاماً على عودة الديمقراطية، تصعب مهمة الوصول إلى البيانات العامة. كما أن بيانات الإنفاق العام تنقل القليل إلى أغلبية الناس الذين ليسوا ضليعين في المحاسبة المالية والحسابات المعقدة.

ومع ظهور الأجهزة المحمولة وتزايد عدد مستخدمي الإنترنت في نيجيريا، ومع BudgIT، رأينا فرصة هائلة لاستخدام تقنيات التوضيح المرئي للبيانات للشرح وإشراك الناس في الإنفاق العام. ولفعل ذلك اضطررنا إلى إشراك المستخدمين من جميع المنصات وإلى التواصل مع المواطنين من خلال المنظمات غير الحكومية. يتعلق هذا المشروع بجعل البيانات العامة أمراً اجتماعياً وبناء شبكة واسعة تطالب بالتغيير.

للنجاح في إشراك المستخدمين، علينا أن نفهم ما الذي يريدونه. ما الذي يهتم به المواطن النيجيري؟ أين يشعر بوجود فجوة في المعلومات؟ كيف يمكننا أن نجعل البيانات ذات صلة بحياته؟ إن الهدف المباشر لمشروع BudgIT هو النيجيري العادي المتعلم الذي يتصل بالمنتديات على الإنترنت والإعلام الاجتماعي. وحتى نحظى بالانتباه المحدود للمستخدمين الغارقين في بحر واسع من الاهتمامات (الألعاب والقراءة والتواصل الاجتماعي)، علينا أن نقدم البيانات بشكل مختصر وموجز. بعد نشر لقطة للبيانات في تغريدة أو رسم بياني، هناك فرصة لمشاركة أكثر استدامة في تجربة أكثر تفاعلية لتزويد المستخدمين بالصورة الكاملة.

عند توضيح البيانات مرئياً، من المهم أن نفهم مدى إلمام مستخدمينا بالبيانات. وعلى الرغم من كونها جميلة وراقية، قد لا تتواصل الرسوم البيانية المعقدة والتطبيقات التفاعلية مع مستخدمينا بشكل مجدٍ حسب تجاربهم السابقة مع تفسير البيانات. والتوضيح المرئي الجيد سيخاطب المستخدم بلغة يمكنه فهمها، ويقدم خبراً يمكن للمستخدم الإحساس به بسهولة.



نموذج 6–27. تطبيق BudgIT (BudgIT نيجيريا)

لقد قمنا بإشراك أكثر من 10,000 نيجيري في الموازنة وتم تصنيفهم إلى ثلاث فئات لضمان تحقيق أفضل قيمة. فيما يلي شرح مختصر لتلك الفئات:

المستخدمون غير المنتظمين

هؤلاء هم المستخدمون الذين يريدون المعلومات ببساطة وبسرعة وما يهمهم هو الحصول على صورة عامة للبيانات وليس التحليلات المُفصلة. يمكننا إشراكهم من خلال التغريدات أو الرسوم البيانية التفاعلية.

المستخدمون النشطون

هؤلاء هم المستخدمون الذين يثيرون النقاشات ويستخدمون البيانات لزيادة معرفتهم بمجال معين أو لتحدي افتراضات تلك البيانات. بالنسبة لهؤلاء، نود أن نوفر آليات لردود الفعل وإمكانية مشاركة الأفكار مع زملائهم من خلال الشبكات الاجتماعية.

الطامعون في البيانات

يريد هؤلاء المستخدمون البيانات الخام من أجل التوضيح المرئي أو التحليل. فنقوم ببساطة بإعطائهم البيانات ليحققوا أغراضهم.

وفي مشروع BudgIT، يعتمد إشراك مستخدمينا على الآتي:

إثارة النقاشات حول المواضيع الرائجة حالياً

يتتبع مشروع BudgIT النقاشات سواء أكانت على الإنترنت أم لم تكن، ويسعى إلى توفير البيانات حول تلك المواضيع. فعلى سبيل المثال، مع إضرابات الوقود في يناير 2012، كان هناك غضبٌ متواصل بين المحتجين الذين ينادون بإعادة الدعم على الوقود وتخفيض الإنفاق العام المسرف وغير الضروري. وقد قام مشروع BudgIT بتتبع النقاشات عبر الإعلام الاجتماعي وبنى – في غضون 36 ساعة من العمل – تطبيقاً يسمح للمواطنين بإعادة تنظيم الموازنة النيجيرية.

آليات جيدة لردود الفعل

نحن نتشارك مع المستخدمين من خلال قنوات النقاش والإعلام الاجتماعي.

ويريد العديد منهم معرفة الأخبار التي تكمن وراء البيانات ويسألوننا عن رأينا. لذا نحرص على أن تشرح إجاباتنا الحقائق فقط وراء البيانات ولا ننحاز إلى آرائنا الشخصية أو السياسية. يجب أن تبقى قنوات ردود الفعل مفتوحة للاستجابة بفعالية إلى التعليقات ولإشراك المستخدمين بشكل خلاق لضمان بقاء المجتمع المبني حول البيانات واستمراره.

اجعلها محلية

بالنسبة لمجموعات البيانات التي تستهدف مجموعة معينة، يهدف مشروع BudgIT إلى جعل المحتوى محلياً والدعوة إلى قناة نقاش تتصل باحتياجات مجموعات معينة من المستخدمين واهتماماتهم. كما نهتم على وجه الخصوص بإشراك المستخدمين في القضايا التي يهتمون بها من خلال الرسائل النصية القصيرة.

بعد إتاحة بيانات الإنفاق على الموقع الإلكتروني yourbudgit.com نتواصل مع المواطنين من خلال العديد من المنظمات غير الحكومية. ونخطط أيضاً لتطوير إطار تشاركي حيث يمكن للمواطنين والمؤسسات الحكومية الاجتماع في مباني البلديات لتحديد العناصر الأساسية التي يجب أن تكون على قمة أولويات الموازنة.

تمت تغطية هذا المشروع في الإعلام المحلي والأجنبي، من موقع -CP Africa (http://bit.ly/cp-africa-budget) إلى شبكة بي. بي. سي. (http:// bbc.in/africa-budget). وقد تولينا مراجعة ميزانية القطاع الأمني للفترة بين 2002 و2011 لـ 'يينكا إيبوكون' الصحفية في وكالة أسوشيتد برس. إن معظم المؤسسات الإعلامية من 'الطامعين في البيانات' ويطلبونها منا ليستخدموها في تقاريرهم. نحن نخطط للمزيد من التعاون مع صحافيين ومؤسسات إعلامية في الأشهر القادمة.

—— أولسيون أونيغبيندي، BudgIT نيجيريا

إشراك الناس في بياناتك



(نموذج 6–28: هل يجب إجراء مزيد من التحقيقات؟ Wired.co.uk)

إن الحصول على رد فعل من جمهورك هو بنفس أهمية نشر البيانات. أنت إنسان، أي أنك سترتكب أخطاء وستفوتك بعض الأمور وستكون لديك فكرة خطأ بين الحين والآخر. وجمهورك يُعتبر واحداً من أكثر مواردك إفادة. فيمكنه التأكد من الحقائق والإشارة إلى بعض الأمور التي قد لا تخطر على بالك.

لكن إشراك الجمهور أمر شائك، فأنت تتعامل مع مجموعة من الناس الذين اعتادوا بعد سنوات من استخدام الإنترنت أن ينتقلوا من موقع إلى آخر تاركين خلفهم تعليقات ساخرة. من المهم جداً أن تبني مستوى للثقة بينك وبين مستخدميك، فيجب أن يعرفوا ما الذي سيحصلون عليه وكيف يمكنهم التفاعل معه وتقديم ردود الفعل، كما يجب أن يعرفوا أنه سيتم الاستماع إليهم وأخذ آرائهم على محمل الجد.

ولكن عليك أولاً أن تفكر في نوع الجمهور الذي لديك أو الذي تود الحصول عليه. جمهور سيعلم ويتعلم من نوع البيانات التي تتعامل معها. فإذا كان من قطاع معين، سيتوجب عليك أن تستكشف وسائل اتصال خاصة بهذا القطاع. هل هناك جهات تجارية يمكنك الاتصال بها وتكون مستعدة لنشر مواردك وعملك إلى جمهور أكبر؟ هل هناك موقع إلكتروني لجماعة معينة أو منتدى يمكنك الاتصال به؟ هل توجد مطبوعات تجارية متخصصة قد تريد أن تنشر تقارير حول بعض الأخبار التي تعثر عليها في البيانات؟

من البديهي أن الإعلام الاجتماعي أداة هامة، على الرغم من أنه يعتمد أيضاً على نوع البيانات التي تتعامل معها. إذا كنت تنظر في إحصائيات شحن عالمية مثلاً، فمن غير المرجح أن تجد مجموعة على فيسبوك أو تويتر تهتم بعملك بشكل خاص. أما إذا كنت تدقق في مؤشرات الفساد حول العالم أو إحصائيات الجرائم المحلية، فمن الأرجح أن يكون هذا مهماً لجمهور أكبر.

بالنسبة لموقع تويتر، فإن أفضل طريقة غالباً ما تكون التواصل مع شخصيات بارزة وشرح أهمية عملك لهم مع إضافة رابط. إذا حالفك الحظ، سيقومون بإعادة نشر تغريدتك إلى قرائهم أيضاً. وهذه طريقة رائعة لزيادة عرض عملك بأقل جهد، ولكن لا تضايق الناس!

بمجرد وصول الناس إلى الصفحة، عليك أن تفكر في كيفية تفاعل جمهورك مع عملك. بالطبع قد يقرأون الخبر الذي كتبته وينظرون إلى الرسوم البيانية أو الخرائط، ولكن إعطاء مستخدميك منفذاً للاستجابة هو أمر في غاية الأهمية ومن المرجح أن يعطيك، أكثر من أي شيء آخر، رؤية أفضل للموضوع الذي تكتب عنه مما سينبهك إلى المزيد من العمل على الموضوع في المستقبل وفي اي اتجاه ينبغي أن تولي الاهتمام والتركيز.

أولاً، من البديهي أن عليك نشر البيانات الخام إلى جانب مقالاتك. إما أن تستضيف البيانات في نص عادي مقسم بالفواصل، وإما أن تستضيفها في خدمات الطرف الثالث مثل وثائق جوجل. وهكذا تكون هناك نسخة واحدة من البيانات، ويمكنك تحديثها وتصحيحها عند الضرورة إذا وجدت بها أخطاء تحتاج إلى التصحيح لاحقاً، ومن الأفضل فعل كلا الأمرين. ساعد الناس بقدر الإمكان بتسهيل حصولهم على موادك الخام.

ثم أبدأ في التفكير فيما إذا كانت هناك وسائل أخرى لحث الجمهور على التفاعل. راقب أي الأجزاء من مجموعات بياناتك تجذب الانتباه – ومن الأرجح أن تكون الأجزاء الأكثر مشاهدة تحوي أمراً ما فاتتك ملاحظته. فعلى سبيل المثال، قد لا تفكر في النظر إلى إحصائيات الفقر في آيسلندة، ولكن إذا كانت تلك الخانات تجذب الكثير من الانتباه، فقد يكون بها ما يستحق التحقق منه.

فكر أيضاً فيما يتخطى مساحة التعليقات. هل تستطيع ربط التعليقات بخانات

معينة في الجدول البياني؟ أو ربطها بمنطقة معينة في الرسم البياني؟ بينما لا تسمح بذلك أغلب أنظمة النشر القابلة للتضمين، يستحق الأمر النظر فيه إذا كنت تصنع شيئًا أكثر تفصيلاً. ولا يمكن التقليل من فوائد هذا الأمر بالنسبة لبياناتك.

تأكد من استطاعة المستخدمين الآخرين أيضاً أن يروا تلك التعليقات، فهي تقريباً بنفس قيمة البيانات الأصلية في الكثير من الحالات. وإذا احتفظت بتلك المعلومات لنفسك، فأنت تحرم جمهورك من الاستفادة منها.

وأخيراً، قد يود الآخرون أن ينشروا رسومهم البيانية وأخبارهم بناء على نفس مصادر البيانات. فكر في أفضل طريقة لربط تلك الأعمال ببعضها البعض. يمكنك أن تستخدم هاشتاج # خاصاً بمجموعة البيانات مثلاً، أو يمكنك مشاركتها مع مجموعات في موقع فليكرFlicker، إذا كانت تصويرية بشكل لافت.

يمكن أيضاً الاستفادة من وسيلة خاصة لمشاركة المعلومات بشكل أكثر سرية. ففي بعض الحالات، قد لا يكون من الآمن مشاركة الناس لمساهماتهم في مجموعة البيانات علنياً، وربما كانوا ببساطة غير مرتاحين لقيامهم بذلك. قد يفضل هؤلاء الناس تقديم معلوماتهم عن طريق البريد الإلكتروني أو حتى مساحة تعليقات مجهولة المصدر.

أهم ما يمكنك فعله ببياناتك هو مشاركتها على أوسع نطاق وبأكبر قدر من الصراحة. عندما تتيح لقرائك التحقق من عملك والعثور على أخطائك وانتقاء الأمور التي قد تكون فاتتك ملاحظتها ستجعل صحافتك وتجربة قرائك أفضل إلى حد بعيد.

—— دنكان جير، موقع wired.co.uk

هذا الكتاب

ولد هذا الكتاب بعد ورشة ضمت عشرات من الصحافيين وخبراء الترميز الحاسوبي وخبراء التوضيحات الغرافية عقدت في لندن عام 2011. ويحتوي على مئات الأمثلة لأخبار احتلت صدارة الصحف والمواقع الإلكترونية تم استخراجها من أكوام المعلومات والأرقام المتاحة في الإنترنت والدفينة في السجلات الرسمية

- وسط الجداول ثمة حقائق خفية، وبالخبرة والصبر يمكن استخراجها.
- صحافة البيانات تنتج أخباراً تمس حياة الناس، قد تكون أخباراً محلية، وقد تكون دولية.
- صحافة البيانات توفر الخبر، وتوفر متابعة له.
- تجارب ومكاسب كبريات الصحف ومحطات التلفزة من صحافة البيانات.
- كيف يتعاون الصحافي والخبير الحاسوبي وخبير الرسوم الغرافية لاستخراج الخبر، ثم التعبير عنه بأوضح صورة.
- كيف يقرأ الصحافي الجداول المعقدة، وكيف يتجنب المزالق الكثيرة التي يسببها غياب بعض الأرقام.
- كيف أثرت ويكيليكس في تفكير الصحافيين الاستقصائيين.
- كيف يتجنب الصحافي الانجراف وراء مؤشرات مضللة، وكيف يظل بعيداً عن التهويل.